DOCUMENT RESUME

ED 139 002 CS 203 386

AUTHOR Haskins, Jack B.
TITLE An Introduction to Advertising Research; A Report from the Communications Research Center.
INSTITUTION Tennessee Univ., Knoxville. Coll. of Communications.
PUB DATE Mar 76
NOTE 249p.; First draft

EDRS PRICE MF-$0.83 HC-$12.71 Plus Postage.
DESCRIPTORS Financial Support; Guides; Higher Education; Management; Mass Media; *Publicize; *Research; *Research Methodology; *Research Tools
IDENTIFIERS *Advertising

ABSTRACT

The purpose of this volume is to present, in nontechnical language, most of the basic concepts of advertising research. Since the volume is intended to be comprehensible to the lay person, discussion does not go too deeply into the technical details of advertising or research methodology. However, used as an introduction and outline to be supplemented by outside readings, the book can be useful as a guide to greater mastery of advertising research practices. An introduction describes advertising and research processes and outlines the most common advertising research plans. Major sections are devoted to research on advertising media, advertising messages, and money and management. Additional facets of advertising research--miscellaneous tools and techniques and the public's attitudes toward advertising in general--are also investigated. (KS)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED139002

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A Report from the Communications Research Center

An Introduction to Advertising Research

by Jack B. Haskins, Ph.D.

(First draft, March, 1976)

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

Jack B. Haskins

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

Communications Research Center
College of Communications
The University of Tennessee, Knoxville 37916

CS 203 336

AUTHOR'S NOTE

This is the first draft of a manuscript which is intended to be eventually published in book form. Being a draft, it is subject to inaccuracies, omissions, typographical errors and over-inclusions which will be corrected in subsequent drafts. The reader is encouraged, therefore, to be alert for possible deficiencies while remaining tolerant of the all-too-human frailties of the writer.

Suggestions for improvement will be welcomed.

All rights are retained by the author.

--Jack B. Haskins, Ph.D
Director
Communications Research Center
College of Communications
The University of Tennessee
Knoxville

March, 1976

FOREWORD

The title of this work, An Introduction to Advertising Research, should provide clues to the author's intentions regarding its purpose and its intended audiences.

Its purpose is to present in plain non-technical language most of the basic concepts of advertising research. It is intended to be readable and comprehensible to the intelligent layman unversed in either the field of advertising or the intricacies of research methodology. In that light, it cannot go too deeply into the technical details of advertising or research methodology. However, used as an introduction and outline to be supplemented by outside readings (to be cited in subsequent drafts), it can be useful as a guide to greater mastery of advertising research practices.

If it is indeed "comprehensible to the intelligent layman," then perhaps it will be useful to these kinds of audiences:

--Advertising practitioners and decision-makers with no great background in research methods who wish to become discerning consumers and users of research without becoming enmeshed in technical details.

--Skilled researchers from outside the advertising field who wish to learn some of the applications of research to advertising as well as the technical variations in technique and method that are unique to the advertising field.

--Students of advertising and marketing who wish to have an overview of the role of research in the total advertising process in order to become discerning consumers and users in the future. At the undergraduate level, additional readings of moderate complexity should be recommended. At the graduate level or as an advanced advertising research text, readings of a more technical nature should be provided.

--JBH, March 1976

CONTENT

FOREWORD iii

PART A - INTRODUCTION

CHAPTER I. THE ADVERTISING PROCESS 1

--Unsystematic Process. 1

--Systematic Process--The Four M's. 2

--Stages of the Advertising Process 3
Vehicle distribution / Vehicle exposure / Advertising exposure / Advertisement perception / Advertisement communication / Behavioral response

--Summary of the Advertising Process. 15

CHAPTER II. THE RESEARCH PROCESS 17

--The Scientific Approach to Advertising Problems. 17

--Advertising Goals and Their Dependent Variables 20

--Steps in Planning a Research Project. 21
Define problem / Review what's known / Decide research methods / Conduct research / Analyze data / Interpret results

--Other Participants in the Research Process. . . 32

CHAPTER III. THE MOST COMMON ADVERTISING RESEARCH PLANS. 35

--The Sample Survey 37
Distinguishing futures / Sampling / Questionnaire construction / Interviewing / Criteria for evaluating the sample survey

--The Survey Panel 49

--The Laboratory Experiment 50
Distinguishing features / Advantages and disadvantages

--The Controlled Field Experiment 52
Example / Variations / CFX test units / Values of CFX

--Attitude and Image Measurement. 61
Definition of attitude and image / Methods of attitude and image measurement / Monadic attitude scales / Relative preference scales / complex scales / Hierarchy scales

--Controlled Observation. 84
Distinguishing characteristics / Uses of controlled observation

--Content Analysis. 87
Distinguishing features / Uses of content analysis

PART B - RESEARCH ON ADVERTISING MEDIA

CHAPTER IV. MEDIA RESEARCH IN GENERAL 93

--Questions Involved in Choosing Media. 94

--Typical Media Research Studies. 94

CHAPTER V. PRINT MEDIA RESEARCH METHODS. 101

--Measuring Distribution of Print Media Vehicles. 102

--Measuring Issue Audiences of Print Media Vehicles 103
Sampling / Question writing / Audience efficiency illustration / Prospect audience efficiency illustration / Accumulative audience / Reach / Frequency / Unduplicated audience

--Measuring Issue Exposures to Print Media Vehicles. 118

--Measuring Advertisement Audiences and Exposures in Print Media Vehicles 122

--Measuring the Print Vehicle's Contribution to Advertisement Perception 125
Reader traffic surveys / Media image studies

CHAPTER VI. BROADCAST MEDIA RESEARCH METHODS. 135

--Mechanical Recordings of Broadcast Reception 137
Instantaneous mechanical measurement

--Dairy Records of Broadcast Reception. 142
All-media diary records

--Coincidental Telephone Interviews or Broadcast Reception. 144

--Other Ways for Measuring Broadcast Reception. . 145
Personal coincidental / Personal roster recall / Personal unaided recall / Telephone unaided recall / Combinations of methods

--Recommended Standards for Broadcast Media Research. 146

--Evaluation of Different Measurement Methods Against Recommended Standards 150

CHAPTER VII. OUTDOOR MEDIA RESEARCH METHODS. 153

PART C - RESEARCH ON ADVERTISING MESSAGES

CHAPTER VIII. MESSAGE RESEARCH IN GENERAL 161

--How NOT To Determine Ad Effectiveness 162

--The Extent of Message Research in Advertising . 164

--Some Questions Involved in Message Research . . 165
Selling concept / Theme / Graphics / Persuasion / Strategy / Body copy / Complete execution

--Typical Message Research Studies. 167

--Pre-Testing, Post-Testing on Pilot Tests. . . . 171

--"What Do You Mean, 'Advertising Effectiveness'?" 173

--Reactions to Product and Reactions to Advertising 174

--Disagreement and Agreement Among Experts. . . . 175
Individual ads / Ad campaigns / Test conditions

CHAPTER IX. PRE-TESTING MESSAGES. 181

--General Pre-Test Guidelines 181
Concept testing

--General Pre-Testing Methods 185
Group discussion methods / Projective methods / Laboratory apparatus methods / Binocular devices / Other general research methods

--Broadcast Message Pre-Testing 191
Commercial form for testing / Different TV pre-testing environments

--Print Message Pre-Testing 196
Ad-rating / Portfolio / Split-run / Content response code / Direct mail / Puzzle game

--Outdoor Message Pre-Testing 201

CHAPTER X. POST-TESTING MESSAGES 203

--Recognition and Recall of Print Ads 204
Recognition and the Starch Score / Controlling recognition errors / Aided recall and the Gallup-Robinson score / Comparison of recognition and aided recall scores

--Recall of Broadcast Ads 209
Telephone survey / Diary

--Conclusion About Recognition and Recall Post-Testing. . 210

--Attitudes and Behavior Toward Product 211
Research design / Awareness / Attitudes, beliefs, buying preference / Inquiries and sales

--Analyzing Post-Test Results to Find Advertising Principles. 217
Combined ratings and content analysis / content analysis of competitive advertising

PART D - RESEARCH ON MONEY AND MANAGEMENT

CHAPTER XI. MONEY RESEARCH: ADVERTISING EXPENDITURE 221

--How NOT To Do It. 221

--A Better Way: Controlled Field Experiment (CFX). . . . 222

--Setting Total Advertising Budget. 223

--Determining the Advertising Spending Pattern. . 228

--Determining Market Allocation of Funds. 230

--Media Allocation Problems 231
Single medium / Constant sum media mix /
Variable sum media mix

--Vehicle Allocation Problems 234

--Other Advertising Expenditures Problems 235

PART E - OTHER FACETS OF ADVERTISING RESEARCH

CHAPTER XII. MISCELLANEOUS TOOLS AND TECHNIQUES. 237

--Operations Research 237

--Statistical Decision Theory 237

--Mathematical Models 238

--Simulation. 239

--Linear Programming. 239

--Experimental Design 240

--Factor Analysis 241

CHAPTER XIII. WHAT THE PUBLIC THINKS ABOUT ADVERTISING. 243

--Attitudes Toward Advertising as an Institution. 244

--Reactions to Specific Advertisements 246

--Summary 248

CHAPTER I. THE ADVERTISING PROCESS

Advertising is most simply defined as an attempt to persuade some people to purchase a product or service by means of messages through paid media.

The first step toward an understanding of advertising research is to acquire some understanding of the process by which advertising decisions are made. Research is a tool that can help make wise advertising decisions.

An advertiser may approach the task of advertising either unsystematically or systematically.

Unsystematic Process

The unsystematic approach--still followed by many companies and advertising agencies to a surprising extent--involves decisions based mostly on personal preferences, traditions, emotional reactions or simply convenience. The most obvious example of the unsystematic approach is the company president who advertises only in those magazines which he and his family and friends like to read, whose advertising expenditure is based on the amount of money left over after other company activities have been taken care of, where advertisements are based on his personal "gut reaction," who insists that his advertising contain boring accounts of company history or his own business philosophy. This approach may be personally very satisfying to the sponsor, but it's unlikely to produce very many sales for his product. The end result may range from, at best, simply a waste of the money spent on advertising, to, at worst, a negative effect which alienates the public and produces fewer sales than if he hadn't advertised at all. And there are documented cases where advertising has had, not just a neutral effect, but a strong negative effect.

Systematic Process--The Four M's

A systematic approach to advertising, on the other hand, would be to pose a series of questions to which objective answers would be sought, questions such as the following:

What is the market for my product?

Which media are most efficient in reaching those people?

Which messages are most effective in persuading those people to choose my product?

How much money must I spend on advertising in order to reach the right people through the right media with the right messages?

The first question--defining the market--is usually a job for marketing research in the pre-advertising stage of product development and deciding on marketing goals. Once the market has been defined, however, it is the job of advertising research to answer the questions about media, messages and money.

Thus, the three main subdivisions within the broad field of advertising research are media research, message or "copy" research, and expenditure or "money" research.

Definitions. These are the four M's of the advertising game--Markets, Media, Messages and Money. Some standard advertising definitions for those terms will be used here. The market consists of the people who are the intended prospective purchasers of a product. The media are the channels of mass communication used in advertising, the major ones being newspapers, magazines, television, radio, outdoor and direct mail. A media vehicle is a specific representative of a medium; the New York Times is a newspaper vehicle, the Readers Digest is a magazine vehicle, KDKA is a radio vehicle, and so on.

A message is any individual advertisement intended for placement in a media vehicle. A campaign is a systematic series of messages.

Stages of the Advertising Process

The systematic approach to advertising may be further clarified by the following model of the advertising process, describing the stages that an advertisement must go through in order to reach its target and have the desired effect.

After a product and its market have been defined, any advertisement placed in any media vehicle must pass through a six-stage obstacle course in order to be successful (ARF, 1961):

- vehicle distribution
- vehicle exposure
- advertisement exposure
- advertisement perception
- advertisement communication
- behavioral response.

In summary, successful advertising involves a chosen media vehicle which is distributed and exposed to the product's potential customers, in which the advertisement is exposed, is perceived, and communicates the intended meaning, culminating in one or more behavioral responses including purchase of the product.

This leads us to a somewhat more precise definition of two major kinds of advertising research.

Media research is concerned with measuring the effectiveness of media vehicles which are responsible for the first three stages of the advertising process: vehicle distribution, vehicle exposure, and advertisement exposure.

Message, or copy, research is concerned with measuring the effectiveness of advertisements which are responsible for advertisement perception and communication.

This breakdown clarifies what the media vehicle is supposed to do for the advertiser, and what the advertising message is supposed to do. Their functions are to a large extent independent of each other, and can be considered and researched separately. To summarize: advertisers generally assume that the advertising vehicle is responsible for vehicle distribution, vehicle exposure, and advertisement exposure; the advertising message is responsible for advertisement perception and communication.

Vehicle distribution. This refers simply to the number of physical units of a given vehicle through which an advertisement can be delivered. "Circulation" is the usual term applied to number of units of vehicle distribution.

In the case of a particular magazine vehicle like the Readers Digest, circulation is the number of copies of a particular issue (or the average number per issue) that are delivered to subscribers and other receivers. Notice that circulation is not equivalent to the number of copies printed, since many of the printed copies may not be distributed and delivered.

In the case of television, vehicle distribution refers to the number of sets tuned in to a particular vehicle or program. Take a local TV station's 6 P.M. news show, for example: its distribution is not the number of TV sets in a station's coverage area, nor even the number of sets turned on, but the number of sets tuned in to the 6 P.M. news program (or some average number for that program).

In the case of outdoor advertising, distribution would correspond to the number of posters carrying the advertising message.

A measure of vehicle distribution is the very roughest measure of a vehicle's value in disseminating advertising. It specifies only the number of inanimate physical units in which the message is carried, but says nothing about the number of people exposed or affected in other ways. However, distribution of the vehicle is a necessary condition for the advertising to have any effect, and in the absence of more precise data--as is sometimes the case--it provides one rough numerical basis for comparing one vehicle with another.

Now let's consider the next stage in an advertisement's journey if it is to have any effect, vehicle exposure.

Vehicle exposure. A magazine copy delivered to a home may be read by several persons, or one person, or by no one. Similarly, a TV set tuned to a particular program may have no viewers in the same room, or it may have several. Obviously, someone must be exposed to a vehicle unit for it to have any value to the advertiser, which explains why vehicle exposure is a better measure of effectiveness than vehicle distribution.

The primary measure of vehicle exposure is the number of people exposed to the vehicle at least once. The people exposed make up the vehicle's "audience," regardless of whether we're talking about newspaper, magazine, TV, radio or any other medium.

It's apparent that a vehicle's audience size can be quite different from it's circulation size. It's possible for audience size to be less than circulation in the case of some vehicles--where there is less than one reader per average magazine copy, or where a large number of tuned in TV sets have no one in the same room. Usually, however, audience size is larger than circulation, and media vehicles can be compared on the basis of audience size.

It's not hard to think of examples where two or more vehicles with the same circulation may be quite different in audience size. A woman's magazine may have only one reader per copy, while a general family-appeal magazine may have three or more readers per copy.

As noted earlier, audience size is a more useful measurement than circulation. It is at the same time, however, more difficult to define and measure, and a great deal of media research is aimed at clarifying precisely what is meant by exposure in terms that will permit comparisons between different vehicles.

In general, a vehicle exposure is a person within sight or sound of a distributed vehicle unit. But to be a useful measurement, it's necessary to be more precise.

Consider a specific magazine copy--let's say the latest issue of Reader's Digest. Which of the following is an "exposure":

- a person who has from a distance seen it lying on a table?
- a person who has glanced idly at the cover?
- a person who has picked up a copy while going through a stack of magazines?
- a person who has riffled through the pages but hasn't read anything?
- a person who has read at least one item completely?
- a person who has spent at least five minutes with the magazine in his hands?
- a person who can correctly identify something from it?

Or consider a specific instalment of a television program--let's say last night's instalment of the 6:30 P.M. NBC news. Assuming that the TV set is on and tuned to the NBC channel, which of the following is an exposure:

- a person who hears snatches of it from another room while doing something else?
- a person who walks through the room?
- a person who is in the room but talking all through the program?
- a person seated in front of the set throughout but who is reading a book?
- a person who watches and listens to the program for one second? one minute? five minutes?
- a person who can remember, unaided, at least one item from program?

Or consider a poster. Which of the following is an exposure:

- a person walking or driving by in either direction?
- a person walking or driving by in a direction facing the advertisement?
- a person driving by with his eyes on the traffic?
- a person whose eyes happen to glance at the billboard?
- a person who is able to recall what was on the billboard?

These may seem like trivial questions until a media buyer is faced with the problem of deciding which of three vehicles is best for advertising a product--a magazine which claims for the average issue 20,000,000 exposures (meaning the number of people who have held it in their hands), a television program which claims for the average evening 20,000,000 exposures (meaning the number of people who have been in the same room at least five minutes), or a poster assemblage which claims 20,000,000 exposures (meaning the number of people who walk or drive by in the average month).

One job of media research is to decide which measure of vehicle exposure is most valid for each media type. Another, and even knottier, is to determine how to measure it.

Then, assuming an exposure to a magazine can be defined and measured and that the same can be done for a television program, how can the buyer be sure that one exposure in a magazine has the same value as an exposure in television?

So far, vehicle exposure has been described as the total number of people exposed to a vehicle unit. Another important factor is the number of exposures per person--two exposures by the same person are probably more valuable than only one. Different vehicles may have the same audience size but differ greatly on number of exposures. If magazine A has 20,000,000 exposures (10 million people exposed twice each) and magazine B has 20,000,000 exposures (20 million people each exposed once), are they of equal value to the advertiser?

Vehicle exposure, then, can refer both to the number of people who are exposed (audience size), and the total number of exposures, which is audience size multiplied by the average number of exposures per person.

Note that nothing has yet been said about the number of people exposed to the advertisement itself, which may be something quite different from the number exposed to the vehicle. This takes us to the next stage of media evaluation and comparison--advertisement exposure.

Advertisement exposure. This means exposure to a specific advertisement in a specific vehicle unit. "Number of advertisement exposures" is a more exact measure of advertising effectiveness than "number of vehicle exposures"; it eliminates those persons who were exposed to the vehicle but not to the advertisement. Vehicle exposure sets an upper limit to advertisement exposure.

The general definition of an advertisement exposure is "one person within sight or sound of the particular advertisement in question." Advertisement exposure can be specified in terms of the number of people exposed, or the total number of exposures, just as in the case of vehicle exposure. The problems of definition in precise measurable terms are even more difficult than for vehicle exposure. How can the researcher determine whether a person saw a particular magazine advertisement, and if so whether he saw it in the particular vehicle under evaluation or somewhere else? Or, assuming a person recalls seeing a particular TV commercial, how can the researcher determine where and when he saw it?

Measures of advertisement exposure provide the most direct index to describe how well a media vehicle confronts people with an advertising message, but says nothing about the quality or effects of the message itself.

In fact, advertisement exposure is a function of how well a vehicle delivers an audience to a space that can be occupied by any advertisement, without considering the quality of the message itself. Looked at in this way, it might be more precisely defined as "the opportunity for advertising perception to take place" and has been called advertisement space exposure.

Advertisement perception. One is on fairly firm ground in talking about vehicle distribution, vehicle exposure, and advertisement exposure; they are relatively easy to describe, define, and count. Advertisement perception, however, is an entirely different kind of measurement, since it deals with a person's response to a message.

Perception is defined as conscious awareness of objects or other data through the senses; in this particular case, perception means awareness of or attention to an advertisement.

Advertisement exposure can be distinguished from advertisement perception by the following diagram:

	Advertisement Exposure	Advertisement Perception
Television	a person in the same room with a TV set turned on, but who doesn't hear or see a commercial when it comes on because of daydreaming, conversation, other activity, etc.	a person who is aware of the sound or picture or both from a TV commercial, even though he may deliberately try to screen it out, as in the case of some annoying commercials.
Magazine	a person idly flipping through the pages of a magazine while thinking of something else but who doesn't see the advertisement even though it passed in front of his eyes.	a person flipping through the pages of a magazine who sees or reads a part of the advertisement.

A person exposed to a newspaper page provides an interesting example of the distinction between exposure and perception. Eye-movement cameras have been used to photograph all the places on a newspaper page over which a person's vision has wandered. Previous tests show that a person's gaze will generally pass over or "scan" almost every item on the page, no matter how small. When asked to go back over the same page with an interviewer and point out what he has seen, however, the person can report seeing only a fraction of those items; some of them he is not at all aware of having seen. In this case, exposure has occurred on all those items over which his gaze passed; perception has occurred only on those items which he is at least aware of having seen. One magazine study found that about two out of three exposures result in measurable perception.

"Aided recall" and "recognition" scores of ads are measures of perception commonly used by the print media.

In practice, there is a very fine line between exposure and perception on the one hand, and between perception and communication on the other and this poses problems of measurement which have not been altogether resolved. Perception is, nevertheless, important to the advertiser because it can provide him with a basic measure of the effect of his advertisement--if he can define it and if he can measure it.

In addition to simple awareness of an advertisement's existence, advertisement perception includes other effects such as interest in, recall or recognition of, knowledge of and attitude toward the advertisement (as distinguished from product or broad awareness, interest, knowledge and attitude, which may be considered functions of advertisement communications).

Thus, one distinction between perception and communication is that the former refers to advertisement responses and the latter refers to product responses.

Advertisement communication. If an advertisement adds to someone's product knowledge, or affects his product attitude, or stimulates a desire to take some sort of product action, then it has communicated to him. Conversely, if someone reports seeing an advertisement, but there is no evidence of increased product knowledge or attitude change or action-arousal, then it has been perceived but it has not communicated.

In the field of psychology, these three kinds of effects are labelled as "cognitive" for knowledge, "affective" for attitude, and "conative" for action-arousal. There are varying degrees of communication within each kind. Obviously, several different measures may be necessary to determine the kind and degree of communication that has occurred. "Awareness" or "salience" is one such measure in the cognitive category.

"Unaided recall" of some aspect of an advertisement is another; this means that something from the advertisement can be remembered, that one or more particular elements of the advertisement had an impact on the memory cells.

It has been demonstrated, however, that awareness and factual recall do not necessarily result in a change of attitude. Consequently, various measures of attitudes toward a product have been devised and used to get at this important kind of communication. Attitude generally refers to the overall liking or disliking for a product, the global reaction to the whole product along a favorable-to-unfavorable continuum. But when one speaks of the feeling about some specific aspect, trait or attribute of a product rather than the whole product, this is referred to as the "product image." In the case of a car, some of its attributes are power, attractiveness, economy, durability, comfort, ease of maintenance, and so on. An advertisement may change some aspect of the product image without necessarily changing the overall attitude--for example, if an automobile advertisement stresses horsepower, the reader may become more convinced that the car is powerful, but if increased power is unimportant to him his attitude will not change. (A subsequent section on attitude measurement deals with this in more detail.)

Even though product knowledge and product attitudes change as the result of an advertisement, this may not be sufficient to produce "action-arousal"--a strong motivation or desire to take some sort of action regarding the product. In the case of a car, action-arousal could take the form of a desire to visit the dealer's showroom, or a desire to take a ride or drive in the car, or a strong intention to purchase that car.

Note that action-arousal is not the same as the action itself--only the desire to take action. It is considered a somewhat stronger advertising effect than attitude change, because attitude change does not necessarily result in the desire to take action. It's considered weaker than behavior, since other things may intervene to prevent the action, or behavior, itself from taking place even though arousal (desire, conation, motivation) may be present.

In summary, then, successful communication may be measured by an increased awareness of or knowledge about a product, or by a change of attitude or image, or in a desire to take some sort of action regarding the product, or in some combination of those responses. There is much disagreement among advertising researchers as to which of these is the best measure of advertising communication. It is probably safest to say that they all measure some aspect of communication effectiveness, that it is not yet known which one is most important or if they're equally important. Logic indicates that attitude change or action-arousal are probably closest to the ultimate criterion, purchasing the product.

Behavioral response. A behavioral response to advertising can take many forms, since it refers to any sort of observable behavior resulting therefrom. The ultimate and most desirable behavior is purchase of the product. There are various intermediate forms of behavior which may be necessary and/or desirable--talking about the product, visiting a retail outlet to see or touch the product, borrowing or otherwise trying the product, and so on. But our concern here is with the ultimate response, a purchase.

The purchase response has two important advantages over such psychological responses as perception and communication. The first is that it measures exactly the effect advertising wants to achieve. One might quarrel with the notion that an advertisement needs to communicate product facts which stick in the consumer's memory, but no one can deny that a purchase of the product is desirable. The second advantage, especially to the advertising researcher, is that purchase behavior is often easier to measure. Behavior is something that can be seen, pointed at and counted, whereas perception and communication cannot be observed inside the consumer's mind and must generally be assumed on the basis of what he says was perceived or communicated.

One difficulty with using purchase response as a measure of advertising effectiveness, however, is that it is hard to disentangle the effect of advertising from the numerous other factors which influence purchase.

Take a single act of purchasing--it is the resultant of such things as product quality, previous experience with the product, availability of a retail outlet, the possession of enough money or credit, the absence of deterring factors such as a veto by another family member or a cut-rate campaign by a competitive product. It's easy to see that a purchase can take place without advertising, or even despite poor advertising (which is worse than none); conversely, despite the best of advertising, the overriding influence of other factors may prevent a sale from taking place.

This illustrates the difficulty of separating the influence of advertising from all the other influences on sales.

There is one way in which advertising's behavioral effects can, however, be isolated, and that is through controlled field experiments; these will be treated in more detail in a later section. Suffice it to say at this point that the controlled field experiment is the only completely acceptable method which can separate the effects of advertising from all the other influences on sales.

Summary of the Advertising Process

This chapter has dealt in some detail with various stages in the advertising process. While not dealing with advertising research per se, this model of the way advertising spreads and influences can be useful in distinguishing the various functions with which research must deal; it sets the stage for the succeeding chapters on various phases of research.

A systematic approach to effective advertising demands...

> ...that a product's market, or potential customers, be defined and described,
> ...that advertising media which are efficient in reaching the product's market be found,
> ...that these media vehicles be distributed to the market,
> ...that the vehicles must be exposed to those persons,
> ...that the advertising within the vehicle be exposed,
> ...that the advertising be perceived,
> ...that some form of communication take place, after which
> ...the goal of a sale may be reached if other influences on sales are not overriding.

Vehicle distribution, vehicle exposure and advertising exposure are generally considered to be the responsibility of the media vehicle in which the advertising is carried. The evaluation and comparison of media vehicles falls within that category of advertising research called media research.

Advertisement perception and communication are generally considered to be the responsibility of the advertising message. The evaluation and testing of advertisements is another category of advertising research called message, or copy, research.

Vehicle distribution, vehicle exposure and advertising exposure are relatively easy to define and less susceptible to verbal distortion, but they represent only intermediate steps toward the advertiser's ultimate purpose.

Advertising perception and communication are difficult to define and not amenable to direct measurement, since generally one must rely on the consumer's verbal report, which may or may not be accurate.

Of all the possible responses to advertising, a purchase has the most to recommend it. A purchase is amenable to direct measurement by the researcher without reliance on verbal reports. It is indisputably relevant to the main purpose of the advertiser. The principal drawback to advertising researchers lies in the heavy influence of non-advertising factors. This difficulty can be surmounted through the use of rigorous research methods, principally through the controlled field experiment. However, a single advertisement is usually not a sufficient amount of communication to bring about a sale, except in special cases; therefore the sales effects of advertising can usually be detailed only for campaigns and not for individual advertisements.

CHAPTER II. THE RESEARCH PROCESS

Just as there is an advertising process, an orderly model of how advertising works, so there is a research process. This process involves both a scientific approach to problem-solving and an orderly procedure of steps to be followed in planning and executing a research project.

The Scientific Approach to Advertising Problems

The creation of advertising is the resultant of two kinds of thinking--the business approach and the creative approach. These approaches are usually housed in two different kinds of people found in advertising--the business mind and the creative mind, only rarely found in the same person. The business mind is concerned with marketing, the practical and tangible aspects of running a business, and making profits. The creative mind is primarily concerned with the artistic, creative and intangible aspects of communication. For many years, advertising has thrived through the combined efforts of these two approaches. The public image of the advertising executive as a sophisticated, persuasive, dynamic, glib and business-like fellow with a satchel full of eye-catching ad layouts is really a composite of the business and creative approaches.

In recent years, however, there has been a growing feeling among advertisers--hard-headed businessmen themselves--that another approach is needed to round out their advertising efforts. As one businessman expressed it, "I know that half my advertising is wasted but I don't know which half." In short, much advertising has been conducted with no "feedback" on its effects--that is with no knowledge of the eventual effectiveness of the effort, no understanding of the general principles of how and why advertising succeeds or fails, no real evidence to demonstrate success or failure.

Such advertising is analogous to a hunter with an expensive high-powered rifle, firing madly at a group of moving targets in the black of night, then moving on complacently without bothering to see if he has actually hit his target, satisfied simply because he has been firing away.

The missing ingredient in advertising in many cases has been the scientific approach. Broadly, the scientific approach simply means an orderly, objective investigation to gather verifiable facts that will add to knowledge, understanding and control over future activity.

Like many other fields of social and behavioral study, advertising is certainly not a science, but it is nevertheless possible to use a scientific approach. The scientific approach--whether it be in chemistry, politics, psychology or advertising--has certain distinguishing characteristics:

The procedures are explicit. The results and the methods for obtaining them are so clearly described that another competent practitioner can read the research report and replicate the procedure if he wishes. The detailed description is minutely honest down to the finest detail of procedure.

The data-collecting is objective. Once a procedure has been formulated, and an investigation started, the researcher is bound to follow the data and report it objectively, whether or not it agrees with his preconceived notion of the outcome. The advertising researcher must fight against perceiving or using only the data which will "prove a point," either his own or some one else's. Quite often, research will come up with the opposite answer than that predicted; this apparently illogical outcome must be perceived and reported according to the revealed evidence.

The findings must be replicable. Other researchers, should they attempt to verify an investigation's results, should arrive at the same conclusion. Intuition, artistic sensitivity, and clinical insight are not sufficient to establish a point of fact.

The approach is systematic and cumulative. The advertising researcher in addition to investigation of an immediate problem, tries to build up whole bodies of knowledge, an organized system of verified facts, generalizations and principles, all based on evidence rather than his hopes or beliefs.

The eventual purposes are explanation, understanding and control. The advertising researcher wants to know why and how something happened, and be able to prove it. Then he can predict and bring about the conditions to achieve his objectives. His objective is the creation of advertising that will sell the product better or more efficiently than any other advertising.

The researcher is both open-minded and skeptical. He is open-minded about accepting the possibility of alternatives to his own pet notions, or to the usual "common sense" answers; in other words, he looks critically at cherished assumptions, thinks of alternatives, and admits them to his thinking as possibly correct. (The hardest alternative for an advertising man to accept is that his advertising is no good; the usual assumption is that there are only two kinds, good and better.) At the same time, the researcher must be extremely skeptical, reluctant to accept anything as fact until it is demonstrated.

The scientific approach to advertising is not merely a high-minded abstraction to be indulged only as a luxury. It is the only practical way to insure against advertising failures and improve the probability of success.

Advertising Goals and Their Dependent Variables

Let's introduce here a technical term which will be useful in talking about advertising research. The term is "dependent variable," and it simply means "the quantity that is measured."

In measuring a person's height, for example, the dependent variable is usually inches. In measuring weight, the dependent variable is usually pounds. In measuring temperature, the dependent variable is usually degrees-fahrenheit. Even in these simple cases, there are other dependent variables that we could use instead: meters instead of inches; kilograms instead of pounds; centigrade-degrees instead of fahrenheit-degrees. All are equally correct, so we must always specify which dependent variables are used.

Similarly, it is not adequate to simply state "Advertisement A was (or was not) effective." To be precise, the measure or criterion or dependent variable involved must be specified. Thus, one can accurately say "Advertisement A was effective "in changing attitudes" or "in increasing consumer awareness" or "in increasing consumer knowledge." Effectiveness is always in terms of some dependent variable which must be specified.

The dependent variable used in measurement should always be directly related to the stated goals of the advertising. Where there are many advertising goals, or where the researcher is uncertain as to which of several possible goals is the proper one, he may include many different dependent variables in a survey; this may include two or three dozen general and specific questions, each designed to measure a different dependent variable.

In short, advertising goal-setting is a prerequisite to measuring effectiveness, and the goals determine which measures of effectiveness or dependent variables will be investigated through research.

Steps in Planning a Research Project

The simplest definition of research is "careful, systematic, patient study and investigation in some field of knowledge, undertaken to establish facts or principles." It is not a haphazard hasty collection of information.

Whether advertising or chemistry or history or any other subject is under study, the research process can be outlined as an orderly procedure, a systematic series of steps which should be followed to get the most useful information possible.

Here is a check-list of the steps that should be followed in planning and executing an advertising research project. This outline is broad enough to be applied to any advertising problem.

1. Define the problem
2. Review what is known about the problem ("Secondary" research)
3. Decide on the research methods to be used ("Primary" research)
4. Conduct the research
5. Analyze the data
6. Interpret and report the results

This may seem like a very precise prescription for action. Within each of these steps, however, a number of decisions must be made, and a great amount of detailed planning is necessary. Let's look at each of these steps in detail.

Define the problem. This means clarification of the problem to end up with one or more specific questions that can be answered empirically. This step is frequently glossed over hurriedly, but it is extremely important; the more time spent discussing and clarifying a problem from a number of different viewpoints, the more useful will be the eventual results.

Who is the person requesting the information? Who will use it once it is produced? What decisions will be based on the results?

A good way to start the problem clarification process is to ask, "What are the objectives of the advertising?" These advertising goals--stated in terms of markets, prospects and intended effects--must be stated explicitly before the researcher can hope to measure the results.

Generally, a problem comes to the researcher from someone else concerned with making advertising decisions, either from the business side or the creative side. It starts as a rather broad and vague request, such as: "Are we getting any results from our television advertising?" It should end up in more precise form such as: "What is the net change in attitudes and sales resulting from the September-June spot-commercial campaign on NBC? Are there any differences in the attitude and sales effects on current users of our own brand, users of other brands, and non-users?" And so on to what may be a list of several specific questions stemming from the same basic problem.

In between the vague request and the eventual specific list of researchable questions, the process of decision, re-thinking and then more discussion at great length is a profitable investment.

The information seeker and the researcher should end this process only when both are mutually agreed that the list of specific researchable questions are indeed the crucial ones to solve the problem. The seeker and the researcher must have the same understanding of what is needed.

Review what is known about the problem. This preliminary step--usually called "secondary research"--is frequently neglected. Many sources of existing knowledge can give general or specific facts and guidance. If the answer to a problem already exists, as is often the case, a great deal of time and money can be saved. There's no point in repeating a research study that has already been done. Even if no specific answers are found, the review of knowledge will help to further define the problem and clarify research procedures to be used.

What are the major sources of existing knowledge? *Company or advertising agency files* form one; if a research department has been in operation for some time, the odds are increased that something useful will be found.

Another source is the *specialized and professional journals* which deal with advertising and marketing. Some of the major ones are:

Journal of Advertising
Journal of Advertising Research
Journal of Marketing Research
Journal of Marketing
Commentary (a British marketing research journal)
Journalism Quarterly
Psychological Abstracts.

These professional journals contain original articles, book reviews, abstracts of previous research, and other material which may be pertinent. Of lesser importance to the researcher, but useful in providing background and leads to other sources are the advertising "news" publications such as Advertising Age. Certainly, no advertising research department is complete without subscriptions to these journals; many city and university libraries also subscribe.

Trade research organizations are another important source. The Advertising Research Foundation (ARF) and the Industrial Advertising Research Institute (IARI) are the major ones.

The ARF is a non-profit organization supported by advertisers, advertising agencies and advertising media. Its basic purpose is to further scientific practices and promote greater objectiveness in advertising and marketing through objective and impartial research. Specifically, ARF offers consultation and guidance on the conduct of research studies; supervises the technical aspect of advertising research projects; evaluates published research studies conducted by others; develops new research methods and techniques, and analyzes existing techniques; establishes research standards and criteria; collects and disseminates advertising and marketing data for the benefit of its members.

All ARF publications are automatically offered to members as they are issued. The regular publications are: The Journal of Advertising Research, a newsletter, and the annual Conference Proceedings. They also issue printed individual reports on supervised studies, consultations, appraisals, analyses of syndicated services, bibliographies, reprints of pertinent material from other sources and speeches. The most useful index to ARF publications is a brochure entitled "Catalog of Publications."

Membership in the ARF is a must for every company or agency involved in advertising; membership at reduced rates is also held by colleges and universities.

The IARI is of particular usefulness to industrial advertisers, those whose advertising is directed toward specialized trade, business and industrial customers rather than the general public; such advertisers customarily use specialized media and different advertising approaches. IARI publishes a variety of material, some of which is useful to the general advertiser.

Two other professional organizations also publish material that is useful: the American Association of Advertising Agencies (AAAA) and the Association of National Advertisers (ANA).

Research conferences are another useful information source. While these may not be available at the particular time needed for a specific problem, nor do they always cover the desired subject matter, they are invaluable to the continuing education of the advertising researcher. Also, most conferences publish reports on what has transpired. Some of the more important ones conducted annually are: the American Association for Public Opinion Research (AAPOR), the Advertising Research Foundation, and the American Marketing Association. In addition to the speeches and published proceedings, such conferences offer an invaluable means of person-to-person contact with other advertising researchers; informal discussion with them often provides information unobtainable in any other way.

Another major source of information is *books and libraries*. There are numerous books on advertising, research methods and marketing research which can provide general and specific guidance to the solution of problems.

There are other sources of information which can be used at this stage of reviewing what is known about a particular problem. But those are the major ones, and it would be foolhardy to ignore them in considering any major research study. Only if this secondary research does not produce the answers should the researcher proceed to "primary research"--the gathering of *new* information.

Decide on research methods. There are an infinite number of ways to conduct an investigation into human behavior. The method of investigation to be used depends on the specific problem, and almost every stated problem is different in some respect from those that have gone before. As a first task in "primary" research, let's consider briefly the seven major methodological aspects in a research investigation: the dependent variable, the response method, the data collection method, sampling, the design, the independent variable, and the research environment.

The dependent variable in a research study is simply the quantity to be measured, as discussed earlier in this chapter. For example, in measuring height, the dependent variable is inches; in weight, pounds; and so on. In advertising studies involving human subjects, the dependent variables depend on the advertising goals or objectives. Some of the major dependent variables in message research are: awareness of the product; knowledge about the product; recognition and recall of advertisements; attitudes and beliefs about the product; behavior concerning the product such as inquiries or sales. In media research studies, some frequently used dependent variables are distribution or circulation, vehicle exposure, advertising exposure, sometimes called vehicle "reach" and vehicle "frequency." All of these dependent variables, and more, are discussed in detail in the chapters on media research and message research.

The instrument used to measure the dependent variable should be valid, reliable, and sensitive. Validity of a measuring instrument means that it measures what it is intended to measure; "inches" is a valid measurement of height but not of body temperature; "number of product facts recalled from an ad" may be a valid measure of learning but not of attitude toward the product. In measuring advertising effectiveness, the purpose of the advertising is the criterion for validity of the dependent variable or measure used; a measure of learning is valid only if the purpose of the advertising is to teach facts. Reliability of a measuring instrument means that it gives consistent results when the same thing is measured on two or more occasions. Where a rubber yardstick measures my height today at 65 inches, and tomorrow at 70 inches, then the measuring instrument is unreliable. Reliability is a necessary but not a sufficient condition for validity; if today I say that two plus two equals five, and tomorrow I say the same thing, I'm being perfectly consistent or reliable, but completely invalid. Precision of a measuring instrument refers to its ability to distinguish between different quantities, the power to discriminate; a ruler marked off at one-foot intervals is not as precise as one marked off at one-inch intervals, that is, it cannot discriminate precisely between fine differences in height.

The data collection method deals with the specific data-gathering process to be used, of which there are several. The major kinds of data collection are: interviews by mail, personal and telephone questionnaire; direct observation of behavior; mechanical recording devices as a substitute for or in addition to personal observation, including films, photographs and recordings; and content analysis of advertising or other data. Each of these data-collection methods is a broad subject of study in itself. A correct decision on which method to use must be based on both "book learning" and actual experience with various methods.

A decision on response method, when human subjects are involved, is based on the previous decisions regarding dependent variables and data collection methods. The major response modes are verbal response (what the respondent says), behavioral response (what the respondent does), and physiological response (involuntary internal changes in the respondent's body).

The sampling of data is another variable. Whether dealing with a study of humans or animals or inanimate objects, it is almost never possible to perform a complete inventory, or census, of the whole population. A sample of those objects, meaning a portion that is representative of the whole population, must be selected. A sample of people, a sample of advertisements, a sample of television stations, a sample of locations--in short, whatever population is involved, some systematic plan for selection of a representative portion must be formulated. The most common application of sampling is found in research involving human subjects; more details on this procedure are given in the following chapter under "The Sample Survey."

A decision on research design deals with the frequency and timing of data collection and the kind of conclusion desired from the study. The "after-only" or "post-test" design involves collection of data on just one occasion; it is a static descriptive study during one slice of time. The "longitudinal" design involves collection of the same data from the same people or sources at several points in time. The "before-and-after" design measures the change that takes place at two points in time, before and after some event, and might be considered a variation of the longitudinal design. The "before-and-after-with-control" design involves data collection at two points in time from two different groups of persons, one group having been exposed to some treatment between measurements (the test group) and one group which has not been exposed to the treatment (the control group); this is actually the classic design for a controlled experiment, the principal and most reliable design for determining cause and effect and the basis of scientific knowledge in many fields of study. The "after-only-with-control" design involves data collection at one point in time (after some planned event) from two groups, a test group and a control group; this also yields cause-and-effect information and is a variation on the classic controlled experiment.

There is one principal distinction in conclusions than can be drawn from the various designs. The first three (after-only, longitudinal, before-and-after) are descriptive studies which can describe the existing situation or tell you what happened, but they cannot determine why they happened--they reveal effects but not the causes of these effects. The latter two designs, both involving a control group, can determine both what happened and why it happened--that is, it determines both cause and effect. More will be said about this extremely important area in the next chapter.

The independent variable is the name given to some treatment deliberately introduced or caused by the researcher, so that its effects can be measured. Sometimes called the "treatment" variable, it is a necessary condition for a controlled experiment. Any quality of advertising that can be manipulated can be an independent variable. For example, if the task is to determine the difference in readership between four-color illustrations, two-color illustrations and black-and-white illustrations, the researcher would construct three ads differing only in those respects; the independent variable in such a case is "use of color." It is the "cause" in a controlled experiment designed to measure cause and effect.

Another research variable is research environment, which may involve both a stimulus environment and a response environment. The stimulus (whatever is used to evoke a response from the respondent) may be a question or an ad or a product or any of a number of objects, depending on the problem at hand. The stimulus may be administered personally or impersonally, from a remote location or in the same location as the respondent, in a laboratory or in the respondent's home or in some other location. One of the most important aspects of the advertising research environment is the nature of exposure--whether exposure to the advertising takes place naturally under real-life conditions of selective exposure, or whether the respondent is forced to pay attention as in an auditorium or studio or classroom. In short, natural exposure vs. forced exposure makes a great deal of difference in the results obtained in communications research. *The most desirable research environment is one in which both exposure and measurement take place in a naturalistic situation without the respondent's awareness that he is the subject of research.*

Conduct the research. After specifing the questions that need answering, the methods of design and data collection have been decided upon and put down in writing in a complete research plan, the researcher is ready to actually execute the research plan. In this phase. the scientific approach is followed--the procedures are explicit, the data-collecting is objective, the plan is followed in detail no matter how the results turn out. Extremely systematic and painstaking attention to detail is necessary.

Analyze the data. This is the exciting phase of a research study, comparable to the thrill of discovery and creation in any field. The preresearch preparation may include dummy tables, requiring only that the blanks be filled in with percentages or scores or other figures. Statistical calculations are performed as necessary, and the figures are rechecked for accuracy. It is frequently possible to check one part of the data against another part, for verification. Where results seem illogical, the entire procedure may be rechecked, to be sure there is no error. Since most advertising research studies are quantitative, the analysis stage will result in a number of statistical tables, summarizing the data in as much detail as necessary. Breakdowns or cross-tabulations of the figures may be necessary. Depending on the magnitude of the job, a computer may be used for extensive calculations; on small jobs, hand-tallying may be sufficient, while others may require a desk calculator or pocket computer.

Interpret and report the results. At this point, after all tables have been prepared, the researcher does a "flashback" to the specific questions formulated in the first step, problem clarification. The principal purpose of the study has been to answer those questions. There are probably more data and more tabulations than is actually needed to answer those specific questions. This requires that the researcher be selective, picking for initial interpretation only those tables which have a direct bearing on the principal specific questions. Additional tabulations can be used to qualify the principal findings.

Regardless of the complexity of the complete study, a short clear understandable report of the principal conclusions must be produced; on the other hand, the report must be complete, reporting all the data, the methods used, and so on. One way of getting around this dilemma is to prepare a "three-stage" report. Part I consists of a simple straightforward one-page report of the principal conclusions--the specific problem, and a short non-technical statement of the principal findings without qualifications. Part II is a somewhat expanded version of Part I, showing the key tables supporting the principal findings, and mentioning the necessary qualifications for those conclusions. It should also contain a brief statement of the research methods used. Part III is a complete technical report, with a full detailed description of methods used at all steps of research planning and all the tables or other data produced by the study.

Parts I ("executive summary") and II should be free of technical jargon and should be understandable by non-research persons. A great deal of advertising research has been wasted because the final reports were too detailed or too technical. Research is useless unless the results are used in decision-making. Decision-makers can't use the results unless they understand what you say. The best safeguard in this respect is to have someone who is unfamiliar with the project read the report before distribution.

Other Participants in the Research Process

These steps involved in a research study have been discussed as though one person or organization were actually doing the whole job. In fact, various consultants and sub-contractors may be brought in at various stages. The amount of outside help depends on the size and skill of the advertiser's or agency's research department, the size and complexity of the research, time available, money available, and other factors.

"Defining the problem" is usually handled by the advertiser and/or the agency, without outside help.

"Secondary research" is usually also an inside job, though university researchers or other specialists may be called in for certain jobs.

In the primary research phases, however, consultants and commercial research firms are frequently used. Data-gathering in particular is almost always farmed out to commercial firms which have standing forces of field interviewers. Data analysis, after the data have been collected, may be sub-contracted to a computer data-processing firm, which will provide all statistical tabulations required. Some commercial research firms will perform both data-gathering and data-analysis; others are specialists for only one of those functions.

While sub-contractors may provide their own interpretation and report of the results, the final interpretations and reports are usually accomplished by the sponsor.

The wide range of commercial research services available are illustrated in two publications--Bradford's Directory of Marketing Research Agencies and the ARF Directory of Research Organization Members.

While all parts of the research process are important, the first step (defining the problem) and the last step (drawing conclusions) are probably most important, and only the sponsoring advertiser or agency can assume final responsibility for their correctness.

CHAPTER III.

THE MOST COMMON ADVERTISING RESEARCH PLANS

Certain kinds of research procedures are more common than others in advertising research. These will be called "advertising research plans," and each involves a particular configuration of design, data collection method, research environment, and other variables.

These "most common" research plans are applicable to a wide variety of advertising problems, and many specific applications are given in the chapters on Media Research and Message Research. At this point, however, a general description of these popular research plans will give a broad understanding of what is to follow.

The most common ones are the *sample survey*, the *survey panel*, the *laboratory experiment*, the *controlled field experiment*, *controlled observation*, *content analysis* and *attitude measurement*.

Each will be discussed in terms of several characteristics: research design, data collection method, response method, research environment, dependent variables, independent variable, and sampling procedure. Some of these features have been discussed in earlier chapters. Response method can stand a little elaboration at this point before detailed description of each plan.

Three response methods. A person's response to any stimulus may be measured in one or more of three ways: *behavioral*, *physiological*, and *verbal*.

When a person deliberately takes some sort of observable physical action, this is known as a *behavioral* response. Behavioral response to advertising can take [illegible]rm of visiting a dealer, mailing a coupon, talking to friends, borrowing a product for a trial, or buying a product. The ultimate criterion of advertising effectiveness is, of course, a purchase--a behavioral response.

When something changes in a person's internal mechanism, this is known as a physiological response--an involuntary reflex action, usually uncontrollable by the person. Physiological response is reflected by heartbeat, sweating, breathing, electrical output of the brain, muscle contraction, and eye pupil size. Since physiological response cannot be determined very accurately by observation, a person must be connected to some sort of laboratory apparatus to get accurate measures. Such changes are often too subtle to be of much use in measuring response to messages. However, all of the above responses have been used in advertising research. One of the most promising of these at present is eye pupil size or "pupillary response"--the pupil dilates (grows larger) when a person is exposed to something unusually pleasing, contracts (grows smaller) when exposed to something unusually displeasing. Brain wave measurement by electroencephalograph (EEG) has also been found related to communications effects.

When a person responds to a stimulus with words, either spoken or written, this is know as a verbal response. Probably the great majority of advertising research takes this form. Mail questionnaires, telephone interviews and personal interviews are all primarily measuring verbal responses. This is the most indirect measurement of the three types, consisting of reports of what a person says he thinks or feels or does. Such reports must always be taken with a grain of salt, since verbal responses are quite controllable by the respondent. A great deal of methodological research has been done to determine what sorts of biases and errors are obtained, and to refine questioning and interviewing techniques to reduce verbal response error. Despite its limitations, the question-and-answer survey is the most widely used data-collection method in marketing and advertising research. In those situations where it is possible to obtain behavioral or physiological response data, they are to be preferred over verbal responses if it is practical to do so.

The Sample Survey

The sample survey--sometimes called a field survey or a poll or simply a survey--is the most frequently used advertising research plan. Basically, it may be described as asking questions of a sample of persons who represent some larger population.

Distinguishing features. In terms of the research methods already discussed, the sample survey may be described as follows:

Research design: one-shot
Data collection method: questionnaire: personal, or mail or telephone
Response method: verbal report
Research environment: natural, real-life (usually at home)
Dependent variables: awareness, knowledge, attitudes, buying preferences (i.e., any verbalizable response)
Independent variable: none
Sampling: usually random

The complete sample survey procedure, after clarifying a problem to be investigated, includes successive stages of planning the sampling, questionnaire construction, data collection and data-processing. Let's discuss here the most crucial technical aspects of these stages in the personal interview sample survey.

Sampling. Sampling is the procedure by which the researcher can determine the characteristics of a large number of people (a population) by analyzing just a few of them (a sample). Sampling is necessary because it would be impossible or financially prohibitive to talk with all the persons in a given population. Fortunately, it has been demonstrated time and time again that a sample--chosen in the proper way--is representative of the population from which it was selected.

First, let's clarify "population." Before sampling, a population must be defined in terms that are meaningful to the specific problem. In advertising, this is frequently a population of prospects for the product. The population in a survey for a soap manufacturer might be "all adults in the U.S." or "all adult women in the U.S." or "all adult married women in the U.S." or "all adult married women who used powdered soap in the last year." The purpose of the survey and the nature of the problem and the kind of product involved, all must be considered in arriving at a very precise verbal description of the target population. Only then can the researcher consider drawing a representative sample from that population.

Frequently, after defining precisely the target population, it is necessary to first select a sample from a larger population, then screen out with appropriate questions the persons to represent the target population. In the case of the soap manufacturer mentioned before, there is no master list of "all adult married women who used powdered soap in the last year"; therefore, the researcher would first select a sample from the "all U.S. households" population, then ask some screening questions to determine age, sex, marital status and soap usage--ending up with a smaller group who are a representative sample of the target population. Only the target population sample are then asked the questions pertinent to the problem at hand.

It is not uncommon for non-researchers to be very skeptical as to whether a small sample can really be representative of a large population. How can a sample of 1,600 persons in an election survey be an accurate reflection of the entire U.S. population of 100 million voting adults? There is not room in this book to go through all the mathematical proof and empirical verifications of sampling theory. For full understanding and acceptance, it is certainly desirable to study the subject further. For the moment, however, you are asked to accept the assumption that scientific sampling of human populations is a valid procedure that has been proved in practice. National election polls, usually accurate within one or two percentage points, provide partial proof.

The first step in sampling is deciding on the sample size, the number of people to be interviewed. The next step is deciding the method of selection, the procedure which will insure getting the right people from the right population.

In the discussions of sample size and method of selection that follow, assume that the aim is to draw a sample from the population defined as "all adults living in U.S. households."

The desirable sample size will be affected by the population homogeneity, the sampling procedure used, practical considerations of time and money and personnel, the desired precision of the results, and the degree of confidence desired in the results.

It may seem strange that size of the population has nothing to do with sample size. A sample of 1,600 persons is just as meaningful for the entire U.S. as it is for a single city, everything else being equal; that is, the sampling error is the same in both cases.

Homogeneity simply refers to the degree to which people are alike on the characteristics being studied. If everyone in a population were exactly alike, one person would adequately represent the whole group. The more the people in a population are alike, the smaller the sample can be. In most studies, however, little information on homogeneity is available and this determinant of sample size is ignored.

Time, money and personnel also affect sample size. Regardless of the other considerations, if there is a limited amount of time, money or people to do the job, sample size must be reduced appropriately.

The number of categories into which the collected data will be analyzed also affects sample size. The more breakdowns used--as by sex or age or education or race--the larger the total sample needed in order to get a meaningful number in each breakdown or cross-tabulation.

Amount of precision required also determines sample size. That is, if it's necessary that results be accurate within five per cent, a larger sample is required than if accuracy within ten per cent is tolerable.

The expected occurrence of a characteristic will affect sample size. This can be known only roughly in advance, but if a split of 50-50 is expected in a two-choice situation (say, a national election), a larger sample is needed than if a 30-70 split is expected.

Finally, the degree of confidence desired in the results affects the size of sample. If 70 per cent confidence is desired, a smaller sample is required than if 95 per cent confidence is needed. Simple tables of sample size requirements are available in most books on survey research and can be mastered with a few minutes of study.

Once the correct sample size is determined, there are several ways of going about the actual sample selection process. A random sample is one in which every individual has an equal chance of selection. When a complete population listing of names or households is available, a systematic random sample can be drawn by designating units at equal intervals throughout the list; for example, to draw a sample of 200 from a listed population of 200,000, you would go down the list choosing each 1,000th name. A simple random sample is obtained by assigning a number to every unit in a population, then selecting the sample members from a table of random numbers. An area-probability sample is drawn by marking off a geographic area into segments (say, census tracts), then choosing from a table of random numbers enough of the numbered segments to give the desired sample size.

The mechanics of drawing a sample are too detailed to be spelled out here. There are non-random procedures (less acceptable) such as quota sampling, availability sampling and others which will not be discussed here. The basic general rule, however, is that no personal biases of the sampler or the interviewer should enter into the selection. Sampling directions should be specified before an interviewer ever goes into the field--specifying exactly where he should go and who he should talk to. Even the most conscientious interviewer will choose the convenient houses and pleasant interviewing situations, and avoid unpleasant or dangerous-looking locations, if left to his own judgment. This would lead to a distinct bias in the survey results. Furthermore, there must be no substitutions in the original sample as drawn; if a person cannot be located the first time, a certain number of callbacks (usually, at least three) must be made. Even then, no substitution is made, since this too would result in some bias. An 80 per cent completion rate for a selected sample is considered good, while 50 per cent completion is considered the minimum rate for unbiased results.

In summary, the principal rule in drawing a sample is to deliberately make as many decisions as possible by chance--flipping a coin, choosing numbers from a random numbers table, and so on. This insures that whatever happens is completely absent of personal biases. Almost anything that is decided subjectively, rather than letting chance decide, is liable to distort the survey results. Understand, "chance selection" is not the same as "haphazard selection"--rather it is a highly deliberate mechanical procedure that is set up to insure a random and therefore representative sample.

Questionnaire construction. The questionnaire is the heart of any sample survey. No matter how carefully the other survey stages are carried out, if the right questions are not asked, or if they're asked in such a way as to distort the results, then it's all meaningless.

Questionnaire writing is an art, rather than a science, and even the most experienced survey researcher follows this general rule; write questions, pre-test them on small groups, revise the questions, then pre-test again and so on until he is satisfied that the questions and the questionnaire format will give the desired information accurately. Five or six pre-tests and revisions are not uncommon; the most experienced question writer is usually surprised at the variety of meanings which respondents get from questions that seemed perfectly clear and direct to him.

Let's consider here four aspects of questionnaire writing--question subject-matter, question structuring, question wording, and question sequence.

"Subject-matter" refers to the kind of information that is needed from the respondent. Most of it can be classified as facts, opinions, knowledge, and reported behavior. A "fact" question asks the respondent for some objective descriptive data about himself--his age, sex, income, marital status, occupation, religion, race and so on; such information is often referred to as "demographic data" or "background data" or "classification data." An "opinion or attitude" question asks the respondent how he feels about something, rather than for objective facts. Generally, a person's feelings for any object--a product, a political candidate, an advertisement--can be determined as for or against, favorable or unfavorable, liking or disliking, the object. All such feelings may not be a simple choice of for or against, and different degrees of feeling may be obtained by appropriate attitude scales, discussed in more detail elsewhere in this chapter.

"Knowledge" questions attempt to find out what the respondent knows, how much he knows, and the source of his knowledge, about various topics. Since answers to knowledge questions depend on memory, sometimes it may be desirable to use aided-recall or recognition questions; knowledge about advertising or about a product or about the mass media may be more easily elicited in this way.

"Behavior" questions ask the respondent to report on some action he has taken--what products he bought and when, which media he has seen or heard recently, and so on. *Reported* behavior is not always the same as *actual* behavior, because of faulty memory or a desire to please the interviewer or other reasons. Consequently, advertising researchers go to great lengths to verify and cross-check such answers, and where possible they turn to actual observation of behavior rather than rely on verbal reports; for example, a "pantry check" of products purchased is sometimes used instead of questions about products purchased.

"Question structuring" refers not to how a question is asked, but to the form that the respondent's reply takes. A completely "unstructured" question allows the respondent to answer in his own words, with the interviewer recording just what is said. The completely "structured" question forces the respondent to give one of several predetermined possible answers. Each has advantages and disadvantages for specific purposes.

An unstructured question--also known as an "open-end" or "free response" question--encourages the respondent to talk freely about the topic at hand. It's especially useful when the researcher doesn't know in advance what kinds of answers to expect or when a great many different kinds of responses are expected, or when he wants the respondent's own thoughts on a subject before asking specific questions, or when he wants to get at underlying motivations. But the unstructured question has some disadvantages; the interviewer must record the answers almost word for word, which may take up considerable time and require considerable space on the questionnaire; if the interviewer abbreviates the answers he may distort the meaning, or the abbreviation may not be clear to the person who later analyzes the answers; since the responses of many people must usually be summarized into categories, the analyst's category system may not accurately reflect what the respondents actually thought.

A structured question presents the respondent with a limited number of fixed answers to choose from. This may range from a simple "yes-no" choice, to a lengthy list of multiple-choice responses. Since the answers are pre-coded with letters or numbers, the interviewer has an easy job recording the answers, and the analyst has an easy job tabulating the responses. The primary disadvantage is that the respondent is forced into a specific answer which may not completely reflect all he thinks about the topic.

Question wording deals not with what is said, but how it is said. People will answer highly intimate personal questions if they're asked in the right way, or they may balk at answering simple impersonal questions if the approach is clumsy. The respondent may misinterpret a question that is poorly worded, and the survey results will be misleading. The best form of insurance against faulty question wording is pre-testing--this preliminary tryout of a questionnaire, and subsequent revision, will catch many wording errors.

The kinds of wording errors are infinite, but three of the major kinds are ambiguity, misperception, and loading. Ambiguity means that a question has more than one meaning; if some respondents interpret a question one way while others interpret it another, the summarized results will be meaningless. Misperception refers to the use of words unfamiliar to the respondent. A loaded question is one that suggests a particular response more than another, or one that is biased in a particular direction.

Other wording problems arise from assuming too much knowledge on the respondent's part, from overly-long and complicated questions that are difficult to follow aurally, from the use of stilted language that sounds unnatural to the respondent, from the use of double negatives, from illogical sentence construction, and many other sources. It cannot be overemphasized that pre-testing, along with knowledge of the survey subject-matter and of the kinds of respondents in the sample, is the best guarantee of correct wording.

Question sequence deals with the order in which questions are asked. Even though individual questions may be properly written, the sequence can affect and distort the responses.

Before a questionnaire goes directly into the crucial questions of a survey, it is often desirable to get the respondent into the right frame of mind through the use of an introduction and some warm-up questions.

The introduction includes the interviewer's very first words to the respondent and everything up until the first question is asked. While the introduction should be written out, the interviewer should try to say it to the respondent in a normal conversational manner, rather than appear to read it from his questionnaire. The purpose of the introduction is to identify the interviewer, his sponsor, and the broad nature of the survey. It should be short and worded in everyday language. It should be straightforward, since a humorous or frivolous approach might convince the respondent that his time would be wasted. It should be neutral and non-partisan, otherwise some respondents might be convinced they're dealing with an enemy. It should be firm and matter-of-fact in assuming that the respondent will cooperate rather than ask for permission to conduct the interview.

The purpose of the warm-up section is to put the respondent at ease with simple easy-to-answer questions. They may be irrelevant to the survey and never be analyzed. Where possible, however, if they can serve the basic purpose and at the same time provide useful information, so much the better.

Past the warm-up stage, the real questions start. One general rule is to start with unstructured and general questions and gradually work up to the structured and highly specific questions. Another rule: if the words in one question suggest answers to another question, then the suggestive question should appear after the other.

Once the main questions in the heart of the questionnaire have been written, the questionnaire should conclude with the demographic or classification questions dealing with facts about the respondent--age, education, occupation and so on.

To summarize, the generally accepted sequence of items within a questionnaire is: introduction, warm-up questions, questions dealing with the principal subject-matter, and finally the demographic questions.

Interviewing. After a questionnaire has been written, pre-tested and rewritten (perhaps several times), the researcher must assemble interviewer materials and brief the interviewers.

Interviewer materials usually include some or all of the following: three-ring notebook, map of the interviewing area and specific addresses or names to call on, the correct number of blank questionnaires identified by case number, extra questionnaires for mistakes and practice, a letter of introduction, income cards or other materials that will be shown to the respondent, instructions and tips for interviewers, expense forms (for time, mileage, etc.) and a writing instrument.

The degree of briefing and training depends on the complexity of the survey and the experience of the interviewers. Before going to work, even trained interviewers need at least one briefing session to familiarize themselves with the questionnaire and the respondent selection procedure. A briefing session may include a demonstration interview by the supervisor followed by questions on procedure. It is often required that each interviewer conduct one or more practice interviews with outsiders, followed by another briefing session with the supervisor.

It would be impossible in this short space to provide a complete "how-to-do-it" procedure for interviewers. Much research has been done and books written on various aspects of interviewing. Some general instructions are: always follow instructions carefully; study the questionnaire for complete familiarity; be neutral, informal, conscientious; read the questions to respondents just as they are written; ask all the questions, and in the order written; record comments completely and accurately; interview only at the place and with the particular person specified; check each questionnaire after the interview to insure completion of every item.

There are certain characteristics common to good interviewers: they're generally interested in people; they're able to inspire confidence and put others at ease; they're inconspicuously but neatly dressed. Mature women are generally highly effective as interviewers, except for special kinds of samples.

Verification of a small proportion of each interviewer's work is a standard practice. The chief interviewer or research supervisor usually conducts this by telephone with a few respondents, asking just a few questions to check on the interviewer's performance. Falsification or error in these verification interviews can lead to scrapping an interviewer's work and dismissal.

Criteria for evaluating the sample survey. Because of the extensive use of sample surveys, and their importance in making advertising decisions, the Advertising Research Foundation has published a list of criteria to be used in conducting and evaluating such studies. Any one of these rules, if violated, may invalidate the results.

1. Under what conditions was the study done? A statement of the method employed should be made available in such detail that the study could be duplicated therefrom.

2. Has the questionnaire been well designed?
3. Has the interviewing been adequately and reliably done?
4. Has the best sampling plan been followed?
5. Has the sampling plan been fully executed?
6. Is the sample large enough?
7. Was there systematic control of editing, coding and tabulation?
8. Is the interpretation forthright and logical?

The ARF brochure on "Criteria for Marketing and Advertising Research" contains a more extensive description of the crucial evaluation points. They parallel some of the points already made in the preceding discussion on the sample survey.

The Survey Panel

The survey panel is simply a sample survey in which the same sample of persons are interviewed on several occasions.

Distinguishing features. The panel survey differs from the sample survey on the following research methods:

Research design: longitudinal
Data collection method: usually mail
Sampling: Random sample of agreeable long-time respondents

The survey panel has certain advantages and certain disadvantages when compared with surveys conducted with different samples of persons. It permits tracing changes in awareness, knowledge, attitudes and reported behavior of the same persons over time. It permits the gathering of certain standard data, such as demographic characteristics, just once rather than in every survey. Some research companies have standardized continuing panels with several participating sponsors; this means that questions on one sponsor's topic is mixed in with questions from other sponsors to reduce research costs.

There are two principal disadvantages. One pertains to the kinds of people in the panel; people who agree to participate in a panel, usually for cash prizes or premiums, are not typical of the whole population in many ways. Another disadvantage is the "conditioning" of the panel members; after a number of surveys, they often become sophisticated or "test-wise" and don't respond in the normal fashion of a naive respondent.

Panel surveys are widely used for certain purposes. It's up to the individual researcher to determine whether this plan is suitable to the problem at hand.

The Laboratory Experiment

The laboratory experiment is the classic plan for conducting controlled studies on human respondents. In this plan, a small number of people can be studied under highly controlled conditions, thus assuring uniformity of the research environment, control of the treatment variables, and various other factors.

Distinguishing features. In terms of the various methods employed in a research plan, the laboratory experiment may be described as follows:

Research design: treatment and control groups, pre-test and post-test
Data collection method: observation, mechanical laboratory apparatus, personal interview
Response method: verbal, behavioral, physiological
Research environment: artificial, unnatural exposure conditions
Dependent variables: all verbal and physiological measures, some behavioral measures

Independent variables: controllable as desired
Sampling: non-random, convenient, available subjects,
small sample. No claim to being representative
of a population.

Advantages and disadvantages. The principle advantage of the laboratory experiment is its ability to determine cause-and-effect relationships between an advertising treatment (cause) and an immediate response (effect). This is achieved through control over the treatment variable, use of a control group, and the research environment. A great variety of data-collection techniques and response methods may be employed.

The principal disadvantages are an unnatural research environment and a non-representative sample. The research environment usually creates forced attention to the treatment variable, often from a highly authoritative person under classroom-like conditions where the respondent is motivated to do his "best," a situation quite different from the normal advertising exposure situation. The subject is aware of measurement procedures which tend to be obtrusive and reactive.

Previous research has shown that laboratory experiments tend to give different results than sample surveys for a variety of reasons; one difference is that sample surveys tend to show either extremely small or no changes in attitude as a result of exposure to advertising, while laboratory experiments tend to show significant changes. There is a strong suspicion that laboratory experimental conditions may artifically inflate the effects of communication.

Nevertheless, laboratory experimentation has produced most of the evidence regarding communications causes and effects. And it is the only method suitable for tightly controlling certain advertising message variables which would be unmeasurable otherwise.

The Controlled Field Experiment

The controlled field experiment (CFX) is probably the most sophisticated and most valid research plan used in advertising research. Not only can it capture the advantages of the two most widely used plans--the sample survey and the laboratory experiment--but it avoids their disadvantages. It can measure cause-and-effect under natural conditions of exposure without the subject's awareness that his reactions are being measured. It is also expensive, time-consuming and requires the most rigorous design and execution of research planning.

It lends itself to measurement of the most important questions in advertising research. How much money should be spent on advertising, and how should it be spent? Which media or combinations of media should be used, and how? Which messages or campaigns are most effective? Results can be measured in terms of behavioral response--notably sales--or verbal response, or both.

The controlled field experiment--or CFX--will be described here in some detail because of its importance and value.

Each of the words in the name of this plan are keys to its distinguishing features.

Controlled means that the researcher controls media placement, message production and distribution, and manipulates the form the advertising takes. For example, if the purpose is to measure the difference in effectiveness of two campaign themes, A and B, the researcher produces the two campaigns to exact specification, then insures that one group of respondents is exposed only to Campaign A and the other is exposed only to Campaign B.

Field refers to the specification that respondents receive the message "in the field" under naturalistic conditions rather than in a laboratory or other artificial situation. They receive it under real-life conditions of selective exposure, without awareness of receiving anything special, or of being involved in a research project at all. Consequently they pay natural attention as they normally do and also respond naturally. Similarly, there is often no awareness that responses will be recorded; behavior and particularly purchasing behavior can be measured through sales registration or auditing records. Under no condition should subjects make a conscious connection between the advertising received, their subsequent behavior, and the measuring process.

Experiment means that the study contains the essential elements of a scientific experiment--one or more treatment groups, a control group and randomization of subjects to treatments. In the simplest form, there is one treatment group and one control group of respondents; the treatment group receives the test message, the control group does not. Measuring the responses of a control group reveals what would have happened without the advertising; the net difference in the responses of the treatment and control groups reveals the incremental effects of the test advertising over and above everything else that has transpired in the test period.

Example: testing campaign themes by CFX. The following simple example will illustrate its principal distinguishing characteristics.

Assume that the advertiser has a new brand of toothpaste called "Sparkle-Dent." It has been on the market for a short while, has received some publicity but no extensive advertising. Two distinctive features have been established in product research--it actually gets teeth white and it really stops bad breath. To conduct a nationwide newspaper advertising campaign, which advertising theme should be used in a newspaper campaign--"white teeth," or "sweet breath"? This question can be answered by a controlled field experiment in a limited area before embarking on a nationwide campaign. The advertising agency is instructed to prepare a month-long newspaper advertising campaign (20 full-page ads, one each day on Monday through Friday for four weeks) for each treatment. One group of prospects will be exposed to the white-teeth campaign and a second group to the sweet-breath campaign. A third group, a control group to receive no advertising of the new toothpaste, is added for reasons that will become clear.

The researcher then selects 12 cities which are similar in a number of crucial respects--population, toothpaste usage, income, one daily newspaper in each, and other population characteristics related to toothpaste buying.

The first step is to randomly divide the cities into three groups, each to receive only one of the three treatments. This might be done in several ways; the simplest is to put each city name on a white card, shuffle the cards thoroughly, then deal them into three piles sight unseen as in a card game.

The next step is to randomly assign the three treatments to the three groups; this may be done with three cards shuffled and dealt blindly to the three groups of cities.

This will provide an arrangement like this:

Campaign A (white-teeth): Cities 2, 5, 11, 6
Campaign B (sweet-breath): Cities 4, 7, 8, 1
Control treatment (no advertising): 3, 9, 12, 10

Space has been purchased in the newspapers of the twelve cities for the test. The next question is, what dependent variables should be measured? This depends on the advertising goal for the product--is the goal increased awareness of the product's existence, increased knowledge of its good features, improved attitudes toward the brand, or simply the ultimate criterion of success, sales? Any or all of these dependent variables could be measured. Since sales measurement is feasible for this product and is the most crucial measure of effects, sales is chosen as the dependent variable in this example (In practice, all of these dependent variables and more might be used.) Sales will be measured both before and after the campaigns, so firm dates for the measurement and advertising are set, and the test is launched.

The time-diagram below illustrates the execution of the test and the results obtained.

	January Sales* before advertising	February Pages of newspaper advertising	March Sales* after advertising
Campaign A	5%	(20)	17%
Campaign B	4	(20)	20
Control	5	(0)	10

*(Proportion of newspaper-subscribing families purchasing one or more tubes of Sparkle-Dent toothpaste.)

The results from the preceding diagram can be reduced to a simple statistical table and the results analyzed as follows.

	Sales			Net Change due to Advertising
Campaign	Before	After	Change	
A	5%	17%	+12	+ 7
B	4	20	+16	+11
Control	5	10	+ 5	--

Note the "before advertising" column. The different city-groups are approximately but not exactly equal in Sparkle-Dent sales, a random variation encountered almost anytime that samples are used. Statistics tables can reveal just how much random variation is tolerable within the bounds of chance. Why the five per cent sales level before advertising? Because the product has been available for awhile, though not widely advertised; only rarely does any product measure "zero" sales.

Now note the results from the "control" treatment. Sales increased from five per cent in January to ten per cent in March even though these cities received no test advertising. This illustrates the real reason for a control group; it reveals that many other things were happening in the toothpaste market apart from our test advertising. Shoppers were seeing it on the shelves and making impulse purchases; word-of-mouth influence was operating; possibly other promotion or publicity was in force, since in a large company it's almost never possible to completely blackout all promotional activities once they're started. The major point, and it is a very important one, is this: if there had been no control group, it would be impossible to determine the magnitude of all the influences on sales apart from the test newspaper advertising. As it is, the changes in the control group are merely subtracted from the changes in the other treatment groups to provide a true measure of the treatment influence. To repeat a basic principle that is essential in any form of research: one can never determine accurately the effects of any treatment unless there is a control group to indicate what would have occurred without the treatment. It's as true in advertising as it is in drug research, where a "placebo" drug is always given to a control group.

The example above is the best way to refute the argument still heard frequently in advertising circles that "you can't really measure the effects of advertising in the marketplace because you can't separate it from all the other marketing forces." It's true that you cannot *control or eliminate* all the non-advertising influences on sales--such as product quality, free publicity, salesmen, word-of-mouth, a reservoir of good will, etc. But the researcher can hold all those other forces *constant* and allow for their effects by the use of two things: randomization of treatments and the use of a control group. The use of the controlled experimental design does give a true measure of the *incremental* effects of any particular advertising treatment over and above other influences that may be operating. All experimentation--whether in psychology, medicine, or in any other field--measures *incremental* effects of the test treatment.

To conclude this analysis of the newspaper advertising experiment, the previous summary is reduced to a still smaller table:

Advertising Treatment	Net Change in Sales Due to Advertising*
Campaign A: White-Teeth	plus 7%
Campaign B: Sweet-breath	plus 11%

*(after subtracting control group change)

It's obvious that campaign B is more effective in selling Sparkle-Dent toothpaste at this time. If, in addition to sales measurement, the researcher had simultaneously conducted sample surveys on awareness, information change, attitude change, and other consumer responses, he could obtain diagnostic data to clarify the reasons why the campaigns operated as they did. In short, by combining verbal response data with the behavioral response (sales), the researcher can determine both what happened and possibly why it happened.

Other CFX variations. The preceding example is just one application of this highly flexible research plan.

The independent or treatment variable in the example was "campaign themes." What other independent variables could be used? To determine the optimum level of advertising expenditures, the following treatments might be used: 50 per cent of normal, normal (control treatment), 200 per cent, 300 per cent and so on; several national advertisers have conducted experiments of this sort. Or the independent variable might be "different media" with treatments as follows: $100,000 in radio vs. the same amount in television vs. the same amount in newspapers vs. the same amount in direct mail, and so on. Another independent variable might be "market size" in which one would set up separate treatments for large cities vs. medium-sized cities vs. small cities. One might also experiment with vehicle treatment: 20 full pages vs. 40 half-pages vs. 80 quarter-pages; vehicle expenditure by contrasting effects of 40 pages vs. 30 pages vs. 20 pages vs. 10 pages. In short, the independent variable in controlled field experiments is limited only by the number of message, media, money and market variations or "causes" that raise questions in the mind of the advertiser or his agency. The examples above do not nearly exhaust the possibilities.

The dependent variables and data collection methods used in controlled field experiments are also quite flexible. In the preceding example, sales was chosen; sales might be measured by store addits, pantry inventories, sales registration data or other means. Other behavioral dependent variables that might be used are: visits to retail dealers, coupon returns, requests for demonstrations and so on. Sample surveys might be employed to measure such communication effects as awareness, knowledge, attitude, buying preference and so on. In short, almost any sort of dependent variable (except possibly physiological responses) can be measured in a controlled field experiment; the nature of the advertised product, product uses and advertising goals should be the principal determining factors as to which of the many dependent variables should be chosen for measurement.

Experimental designs of other sorts are possible in controlled field experiments. The previous example was the "classic" controlled experiment in which only one independent variable is used--the before-and-after-with-control design.

The basic concepts of other designs are relatively simple. In the *factorial design*, the effects of two or more independent variables and their interactions with each other can be determined; for example, different levels of expenditure in different media has been the subject of experimentation by Ford and DuPont. The *latin square* design permits evaluation of the influence of three different variables at once, but cannot measure interaction between those variables; it is simpler but less informative than the factorial design. There are many other sorts of experimental designs--crossover, randomized blocks, and so on. The point is, controlled field experiments lend themselves to any sort of experimental design to measure either simple or complex effects.

CFX test units. Controlled field experiments may also be classified according to the test *unit* which is manipulated. Three of these types are the "test market" technique, the "split-run" technique, and the "checkerboard" technique.

The test-market technique is illustrated by the example used earlier. A group of cities or markets are selected, then the individual markets are randomly assigned to the various treatments. The test unit which was randomized is "markets"--hence the name, test market. The test markets chosen as units are usually non-contiguous, isolated from each other and free from outside media influence.

The split-run technique uses individuals or individual households as the treatment units. Starting from a listing of individuals, in a two-treatment experiment, every other person is designated--AB AB AB AB AB--to form two randomized groups, A and B. Where there are three treatments, every third person is designated--ABC ABC ABC--to form three randomized groups, A, B, and C; and so on for any number of treatment groups. The split-run technique is frequently used in magazine and newspaper advertising research, where a complete listing of subscribers is available for a particular vehicle. It is also useful in direct-mail research, where mailing lists of particular kinds of individuals are available; for example, an automobile advertiser might experiment with two or more treatments among Ford owners, among Chevrolet owners, among ministers, among teachers, or any other group for which a listing is available. It has even been used in television, in those communities where a split-cable feed to alternate households from a community antenna is available.

The checkerboard technique uses areas within regions as the treatment units. A city or other market area can be laid out in grids or census tracts like a checkerboard, and these tracts can be randomly assigned to various treatments, one treatment per tract. The Milwaukee *Journal* pioneered in developing a continuing research facility using the checkerboard technique.

All three of these techniques are amenable to advertisers wishing to experiment with two or more advertising treatments. One variation which can be used with any of them--at somewhat less cost--is the "black-out" advertising method. In an ongoing advertising campaign, after a sample of markets or individuals or tracts has been randomly divided into two groups, the advertiser simply withholds or "blacks-out" the advertising from one group which thereby becomes a control group. Sometimes the savings in media costs may offset the cost of the research.

Values of CFX. The controlled field experiment is the only advertising research plan which can measure cause-and-effect relationships, between advertising and its effects, in a real life situation. It is a relatively new technique and not many advertisers have used it extensively; the number of users is growing, however, and it is significant that some of the biggest and most sophisticated advertisers--Ford, DuPont, American Oil, and Scott Paper, among others--have been quick to adopt it. One important point to remember is that controlled field experiments can be used, in certain forms, by small advertisers as well as large ones. More will be heard about it in the years to come.

Attitude and Image Measurement

Attitude and image measurement is actually just one ingredient that may be used in many of the research plans described and is not, strictly speaking, itself a complete research plan. But because of its widespread usage for all kinds of advertising research problems, it is included here along with the other complete plans.

Attitude and image measurement is used very frequently as an ingredient of sample surveys, laboratory experiments, and controlled field experiments.

Definition of attitude and image. The simplest definition of attitude is "a mental or neural predisposition to respond favorably or unfavorably toward some particular object." For most things with which a person is familiar, he has developed feelings of liking or disliking, for or against, favor or disfavor, in varying degrees. The simplest way to illustrate this is to have you rate your feelings toward the common objects (referents) in Figure 1.

Most people have no trouble assigning an attitude rating to any of the objects above, whether favorable, unfavorable or neutral.

This exercise also illustrates that people can have attitudes toward physical objects (castor oil), toward people, toward abstract concepts (communism), toward organizations (Ford Motor Co.), toward consumer products (Crest), toward product advertising, toward a whole advertising medium (television) or toward a particular media vehicle (Reader's Digest).

The attitude scale example reveals both the direction and intensity of attitude but not the underlying reasons. For example, one's attitude toward Volkswagen cars may be a resultant of many forces--previous experience with the product, its relative preformance compared with other brands, visual impressions, the reactions of friends and family, early childhood impressions, and possibly advertising and other forms of mass communication. It should be noted that advertising is only one of the many influences on an attitude toward a product.

Figure 1

ATTITUDE QUESTIONNAIRE

	Indicate your feeling toward each object by checking one response on each line below.				
ATTITUDE OBJECT (REFERENT)	Strongly like	Mildly like	Neither like nor dislike	Mildly dislike	Strongly dislike
"Hard-rock" music	____	____	____	____	____
Communism	____	____	____	____	____
Spinach	____	____	____	____	____
Chocolate ice cream	____	____	____	____	____
Dogs	____	____	____	____	____
Sunsets	____	____	____	____	____
Crest toothpaste	____	____	____	____	____
Country-western music	____	____	____	____	____
Volkswagon cars	____	____	____	____	____
The U.S. President	____	____	____	____	____
Reader's Digest	____	____	____	____	____
The Ford Motor Co.	____	____	____	____	____
Television	____	____	____	____	____
Caster Oil	____	____	____	____	____

The term image refers to beliefs, reputation or "perceived product attributes." The term is commonly used in many ways--the image of a political candidate, corporation image, product image and media image are only a few. The primary concerns in advertising research are with the latter three.

A brand image is a set of beliefs about the attributes or characteristics of a particular brand of product. It's the "mental description" or profile of a brand with a number of characteristics. In the case of a car, for example, the salient characteristics on which the image is based might be: pickup and getaway power, initial cost, maintenance and service requirements, appearance and styling, safety, riding comfort, trade-in value, durability. (Those are some of the actual image traits used by one major automobile manufacturer in its continuing consumer surveys.) Figure 2 shows how this is measured.

A person's rating of a car on any particular image trait may or may not be related to his attitude toward the car; however, by analyzing brand image data, it's possible to get some ideas (i.e., hypotheses) about the underlying reasons for favorable and unfavorable attitudes. This can be illustrated by analyzing the attitudes and images of a particular car as determined by a consumer survey.

In this example, assume that a consumer survey of 1000 car-buying prospects reveals that attitudes toward Volkswagen are distributed as follows:

Strongly like	(210)	21%
Mildly like	(120)	12%
Neutral	(360)	36%
Mildly dislike	(180)	18%
Strongly dislike	(130)	13%
TOTAL	(1000)	100%

Figure 2

CAR IMAGE QUESTIONNAIRE

Compared with other cars you might consider buying, how would you rate Volkswagen on the qualities below?	(Check one space each line) Above average	 Average	 Below average
Good pickup and getaway?	____	____	____
Low initial cost?	____	____	____
Attractive appearance?	____	____	____
Long-time durability?	____	____	____
Economical operation?	____	____	____
Ease of handling?	____	____	____
Smooth ride?	____	____	____
High trade-in value?	____	____	____
Trouble-free maintenance?	____	____	____
Right size for my needs?	____	____	____

To determine some underlying reasons for liking the Volkswagen, the researcher can break out the image data for just the 210 people who "strongly like" the Volkswagen. Table I shows the results of this cross-tabulation.

TABLE I

IMAGE PROFILE ASSOCIATED WITH FAVORABLE VW ATTITUDE

Volkswagen Trait	Image Rating Distribution Among 210 "Favorable" People			"Above"-"Below" Difference
	Above Average	Average	Below Average	
Good pickup and getaway	20%	55%	25%	- 5%
Low initial cost	70	15	15	+55
Attractive appearance	10	30	60	-50
Long-time durability	40	30	30	-10
Economical operation	80	10	10	+70
Ease of handling	60	10	30	+30
Smooth ride	30	20	50	-20
High trade-in value	20	65	15	+ 5
Trouble-free maintenance	40	20	40	0
Right size for needs	60	30	10	+50

The "image profile" table shows that, among people with a favorable attitude towards Volkswagen, 20 per cent rated it above average on "pickup and getaway" while 25 per cent rated it below average; the difference of minus five per cent indicates that "pickup and getaway" is a mildly negative feature of the Volkswagen. But on "low initial cost," the difference of plus 55 per cent shows this is a highly positive feature. "Attractive appearance," on the other hand, has a difference of minus 50 per cent, revealing this to be a highly negative feature of the Volkswagen.

In summarizing this image data, we may conclude that some "reasons" for a favorable attitude toward Volkswagen are, in order, "economical operation" (+70), "low initial cost" (+55), "right size for needs" (+50), and "ease of handling" (+30). Furthermore, Volkswagen is liked despite some negative features such as "appearance" (-50) and "ride" (-20); these may be called "unimportant reasons" for this particular group of people. Apparently, such features as "pickup and getaway," durability, trade-in value and trouble-free maintenance are of neutral value.

Those are the "reasons" for liking the Volkswagen. A similar image analysis could be performed on the 130 people with unfavorable attitudes; by looking at their strongly negative image ratings, their "reasons" for disliking Volkswagen could be determined.

This example of analyzing attitude and image ratings is only one way of "massaging the data." It is relatively more simple (but possibly less revealing) than such complex statistical techniques as factor analysis, discriminant analysis, multiple regression analysis and the like, for which the use of a computer would be necessary.

An image profile consisting of many beliefs about a brand will often show a "halo effect"--that is, a positive attitude toward a brand would produce many favorable beliefs about the product image.

The definitions of attitude and image are summarized as follows.

Attitude is an overall favorable or unfavorable feeling toward a referent object (product, advertising medium, company, etc.). The image of that object is a set of beliefs about its characteristics. A person can perceive an infinite number of characteristics which go to make up on object's image, but he is said to have only one attitude toward it. An analysis of the image or beliefs associated with an attitude can be useful in determining the reasons why the attitude exists--negative aspects to be corrected and overcome, positive aspects to be capitalized upon. Such analysis is particularly valuable in determining "selling ideas" or "propositions" to be used in advertising. It also is useful in product improvement, since no amount of advertising can hope to overcome an unfavorable belief that is based on a truly bad characteristic of the product itself.

Methods of attitude and image measurement. A brand's attitude and image can be measured by exactly the same methods. All the techniques that are described in the remainder of this section can be applied to both attitude and image measurement. Consumer surveys usually incorporate attitude and image measurement in the same questionnaire; since on the surface they look the same, this often leads to confusion in analyzing the results.

The remainder of this section will discuss attitude measurement only, for simplicity. But keep in mind that these are also tools for image measurement.

The three major ways of expressing an attitude--verbal, physiological, and behavioral responses--have already been described in some detail.

The verbal response, as expressed in sample surveys--by personal interview, telephone, or mail questionnaire--is the most frequent method of attitude measurement in advertising research, and will be discussed in some detail here. (The chapter on message pre-testing will describe physiological attitude research techniques in more detail.)

There are many, many different techniques for the measurement of conscious verbally-expressed attitudes in surveys. These are separated below into several general types--monadic, relative preference, complex, and hierarchy scales.

Monadic attitude scales. Monadic refers to an absolute rating of one object at a time. For example, in measuring the height of two people, John and Bill, monadic measurement requires independent measurement of John's height (say, 73 inches) and Bill's height (say 70 inches). To observe only that "John is taller than Bill" is relative measurement, to be discussed later. Several monadic scales are described below.

The checklist is simply a listing of objects with the instruction to check those that are liked or agreed with. Here's an example:

> "Here's a list of different toothpaste brands. Put a check-mark beside those you like. You may check as few or as many as you like."

Toothpaste brand (attitude object)	Check those you like
Colgate	( )
Sparkle-Dent	( )
Crest	( )

The dichotomous choice requires choosing one of two responses, like or dislike, for each of the objects listed.

Example: "Here's a list of different toothpaste brands. Beside each brand, indicate whether you like or dislike it, by checking the appropriate box. Be sure you put one checkmark for each brand."

Toothpaste brand	Check one	
	LIKE	DISLIKE
Colgate	( )	( )
Sparkle-Dent	( )	( )
Crest	( )	( )

The multiple choice requires choosing one of three or more categorized response expressing varying degrees of attitude. This type is frequently called "a verbal rating scale."

> "Here's a list of different toothpaste brands. Beside each brand, indicate the degree to which you like or dislike it, by checking the appropriate box. Be sure you put one checkmark beside each brand."

	CHECK ONE FOR EACH BRAND				
Brand	Strongly like	Like Mildly	Neither like nor dislike	Dislike Mildly	Strongly Dislike
Colgate........	____	____	____	____	____
Sparkle-Dent...	____	____	____	____	____
Crest..........	____	____	____	____	____

The multiple-choice rating scale can use many other descriptive terms and any number of degrees of attitude. One common usage is simply: Excellent, very good, good, fair, poor. A nine-step verbal scale might take the form: Best of all, like extremely, like quite a bit, like moderately, neutral, dislike moderately, dislike extremely, despise.

The well-known "Likert scale" is named after a prominent psychologist, Rensis Likert, and is widely used in basic communications research; to a series of controversial statements (rather than objects), the subject checks one of these: Strongly agree, agree, neither agree nor disagree, disagree, strongly disagree.

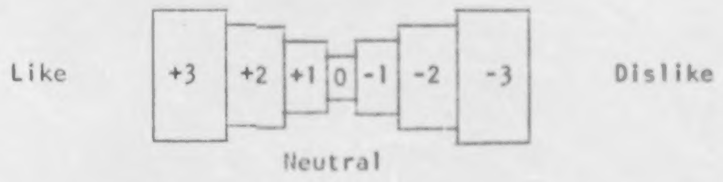
The semantic differential is another widely used technique in advertising. It is really an "expanded dichotomy" with seven steps to allow for varying degrees of feeling. The results from several different scales may be combined to arrive at an overall attitude score.

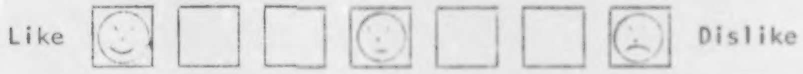
Example: "Indicate your attitude toward each of the toothpaste brands, by marking all three scales for each brand. Be sure you rate each brand on all three scales."

COLGATE TOOTHPASTE: Pleasant___:___:___:___:___:___:___ Unpleasant
Tastes good___:___:___:___:___:___:___ Tastes bad
Like___:___:___:___:___:___:___ Dislike

SPARKLE-DENT TOOTHPASTE: Pleasant___:___:___:___:___:___:___ Unpleasant
Tastes good___:___:___:___:___:___:___ Tastes bad
Like___:___:___:___:___:___:___ Dislike

Many variations of the semantic differential have been devised for use in advertising research. For example, the varying degrees of attitude might be presented in this fashion:





When numbers are attached to the seven steps of the semantic differential, it becomes a combination of verbal and numberical scaling--the numbers may run from +3 to -3, or from 7 down to 1.

With the semantic differential, as with other scales, the scale may run from like-to-dislike, or from dislike-to-like; where several scales are used for each object, the sequence is usually alternated, to avoid stereotyped responses.

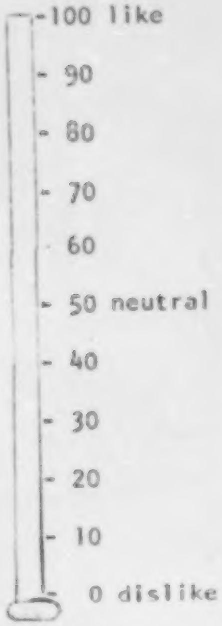
The "calorimeter" is an expanded dichotomy with numbers attached, making it a combination of verbal and numerical scaling.

Example: "For each of the three toothpaste brands named below, indicate your feeling by checking one box in the column below it."

	Colgate toothpaste	Sparkle-Dent toothpaste	Crest toothpaste
Strongly like	(+ 5)	(+ 5)	(+ 5)
	(+ 4)	(+ 4)	(+ 4)
	(+ 3)	(+ 3)	(+ 3)
	(+ 2)	(+ 2)	(+ 2)
	(+ 1)	(+ 1)	(+ 1)
Neutral	(0)	(0)	(0)
	(- 1)	(- 1)	(- 1)
	(- 2)	(- 2)	(- 2)
	(- 3)	(- 3)	(- 3)
	(- 4)	(- 4)	(- 4)
Strongly dislike	(- 5)	(- 5)	(- 5)

The numerical (or quantitative) scale is basically an arrangement of numbers, from low to high, indicating degree of feeling.

Example:

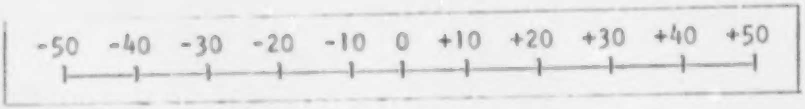


"At the left, you will see a thermometer rating scale, which indicates degrees of liking for a product. The highest possible rating is 100, the lowest possible rating is zero. Beside each of the brands of toothpaste shown at the right, please write in your rating, indicating how much you like it. You can write in any rating from zero to 100 for any brand."

Toothpaste	Write in your rating
Colgate	________
Sparkle-Dent	________
Crest	________

It too can be called an expanded dichotomy, or a combination of verbal and numerical scaling, if verbal "anchor points" are added at the ends of the scale.

The variations on the numerical scale are almost endless. One method of emphasizing negative ratings is to run the scale from minus 100 to plus 100, or minus 50 to plus 50, thus:



Respondents are able to visualize numerical scales in telephone interviews as well as in personal or mail interviews. One ingenious telephone interview usage is the "telephone dial" scale. Respondents are instructed to look at the numbers on the dial, which run 1,2,3,4,5,6,7,8,9. They are then given the name of a product and asked to rate their feelings according to the telephone dial: "1 means strongly dislike, 9 means strongly like" and so on.

This concludes the discussion on monadic scales, with the final thought that these are only a few of the possible variations. The results from all these different kinds of scales are highly correlated with each other. In deciding which of them to use in a particular case, the degree of precision required may be the deciding factor or previous usage may be the deciding factor; everything else being equal, one should use scales that give comparable data to other studies. Ease of use by the respondent is also a very desirable virtue.

Relative preference scales. These attitude ratings involve relative comparisons of one object with another, or with several others.

The paired comparison scale is widely used in advertising and marketing research. No matter how many brands are to be rated, the respondent only makes a comparison of two at the time, until all possible pairs have been rated.

Example: "I would like for you to evaluate three brands of toothpaste for me--Colgate, Sparkle-Dent and Crest.

Tell me which one of each pair you like best ....

Colgate or Sparkle-Dent? __________

Colgate or Crest? __________

Crest or Sparkle-Dent? __________

In summarizing the results of such a survey from a number of people, statistically computed percentages determine the final ranking of the three brands. One problem with this technique lies in the large number of pairs that result when the number of objects to be rated increases. With three brands, there are only 3 pairs; with four brands, 6 pairs; with five brands, 10 pairs; with six brands, 15 pairs, with seven brands, 21 pairs; and so on. It's less time-consuming and less confusing to the respondent to use seven monadic scales than to use 21 paired comparisons.

The ranking or "order-of-merit" method is a way of getting relative preference judgements without going through a large number of paired comparisons. With this method, the respondent is simply asked to rank the objects in order from first to last.

Example: "I would like your evaluation of three brands of toothpaste. Considering these three brands--Colgate, Sparkle-Dent and Crest--which one do you like best? Which is your second choice? Which is your third choice?"

	Mark 1, 2 and 3 for first, second and third choice
Colgate	__________
Sparkle-Dent	__________
Crest	__________

This too can get confusing for the respondent when a large number of objects are to be rated. In personal interviews, this is handled by presenting the respondent with the entire list of objects and letting him put the appropriate rankings beside each one (a "self-administered" question).

One flaw in the ranking method is that no allowance is made for different degrees of feeling between different rankings. For example, assume that a respondent rates Colgate, Sparkle-Dent and Crest as 1,2,3 in that order. His true feeling may be, however, that Colgate is a great deal better than Sparkle-Dent, and Sparkle-Dent is only slightly better than Crest.

The constant-sum rating is one way of indicating the magnitude of difference between different rankings. This involves taking a certain set number of total "points" and asking the respondent to allocate any number of points to each of the different objects being evaluated.

Example: "I'd like for you to evaluate your liking for three different brands of toothpaste. If you had a total of 10 points, how would you divide them up among Colgate, Sparkle-Dent, and Crest? The more points you give, the better you like the brand."

Out of your total of ten points, how many do you give....

	Points
Colgate?	______
Sparkle-Dent?	______
Crest?	______

It's obvious that point allocations of 7, 3 and zero to three brands mean something quite different from 5, 3 and 2--even though the rankings are the same.

One clever variation on this method is to substitute money for points--that is, the respondent is asked to divide up one dollar among the various brands instead of theoretical "points." Sometimes in such a survey, the respondent is given a number of real pennies, or nickels, etc., with the names of the products spread out on cards; the respondent actually places the money on each brand, depending on his evaluation of it. This would seem to be a highly realistic method of obtaining true product attitudes.

Both the monadic and the relative preference scale types are easy to construct without extensive preliminary research.

Complex scales. These scales are not widely used in advertising research for several reasons--they require a great deal of preliminary research, they frequently require cumbersome statistical analysis, and they are not demonstrably more valid than the simpler types already discussed. Three of the better-known ones--Thurstone, Guttman and Q-sort technique--will be discussed here briefly. All three were developed by psychologists for basic attitude research on non-advertising topics.

The Thurstone Scale or "method of equal-appearing intervals" is an 11-step verbal scale consisting of whole statements which have intervals of equal numerical value in between. The respondent chooses the one statement which best describes his attitude toward the test object.

Example: "I'd like to find out how you feel about different brands of toothpaste. From the 11 statements below, please mark the one that comes closest to your own attitude.

First of all, which statement best expresses your feeling toward Colgate toothpaste?

_____ If somebody gave me some, I'd give it away just to get rid of it.
_____ I definitely dislike it.
_____ I would have to force myself to use it.
_____ I would use it if no other was available.
_____ I neither like nor dislike it.
_____ It's OK but there are others I prefer.
_____ It's better than most brands.
_____ I prefer it to all other brands.
_____ It's the only brand I would consider using; perfect.

The scale above is a fictitious one, not developed from basic research. In a real Thurstone scale, the choice of statements is a very tedious process; they're not decided in grab-bag fashion from an armchair. The procedure in devising the scale runs something like this: first, select a large number of statements (sometimes 200 or 300) representing all possible degrees of feeling toward an object. These may come from consumer's free comments, what experts have said, content analysis of ads, anywhere. Each of the, say, 200 statements is typed on a separate card. A number of "expert" judges (researchers, copywriters, marketing people, psychologists, etc.)--as many as 50 to 100--is then selected. Each one is given the deck of 200 statement cards and told to sort them into piles ranging from no. 1 (least favorable) to no. 11 (most favorable)--any number of cards per pile. The average placement of each card by the fifty judges determines its location on the scale. Then 11 of the 200 statements are chosen, based on their statistical average and variability, to represent the 11 degrees of favorability. These are then checked with a separate panel of expert judges, and if they hold up, only then are they ready for use in an attitude survey. The development process is actually much more tedious than described above, but this provides a rough idea.

The Guttman scale or "Cornell method" is theoretically a very sound process, but it too is extremely time-consuming and laborious. It is a "hierarchy" scale in the sense that anyone endorsing or agreeing with a certain scale position also agrees with a lower scale position.

Example: "I'd like to find out how you feel about different brands of toothpaste.

First, let's take Colgate. Beside each of the statements below, put a plus mark beside the highest position on the scale you would endorse."

_____ I'd buy it at $5 per tube.
_____ I'd buy it at $1 per tube.
_____ I'd buy it at 50 cents per tube.
_____ I'd buy it at 10 cents per tube.
_____ I'd take it if it were given to me.

This is one form of "hierarchy" scale, meaning that the various steps form a ladder of successively more favorable attitude. If the highest price checked is the $1 price, the respondent would also buy at all lower prices; if the highest price a respondent was willing to pay was 10 cents, he would not buy at any of the higher prices, but would agree to take it at the lower position. The use of "price willing to pay" in the above example is only one possible way of measuring product attitudes by the Guttman method.

This is known as a "hierarchy" of responses--meaning that agreement with any statement implies agreement with all the less favorable statements and disagreement with all the more favorable statements.

The development of the Guttman scale is so complex that no explanation will be attempted here. The theoretical concepts of unidemensionality and a hierarchy of responses are useful, however, in understanding consumer behavior.

The Q-sort technique of attitude measurement was developed by psychologist William Stephenson. Like the Thurstone method, Q-sort uses statements arranged along an 11-step continuum. It differs, however, in that the judges or respondents have a fixed total number of statements (usually 48 or multiples thereof) and a fixed number prescribed for each of the 11 steps. The purpose of this sorting is to force judgments into a "normal" or "bell-shaped" distribution which facilitates statistical analysis.

When 48 statements are used, the respondent is forced to sort them thus:

	Position	Number of statements
(Most favorable)	No. 1	2
	No. 2	3
	No. 3	5
	No. 4	5
	No. 5	6
	No. 6	6
	No. 7	6
	No. 8	5
	No. 9	5
	No. 10	3
(Most unfavorable)	No. 11	2

Example:

> "Here are 48 cards, each containing a statement about toothpaste. I'd like you to use them to tell what you think about COLGATE toothpaste.
>
> Now here is a sorting board with 11 pockets. The left-hand side of the board is for statements that you think are most descriptive of Colgate, the far right position for statements that are least descriptive of Colgate. Please put two statements on the far left, 3 in the next pocket (etc. through the various positions) and finally 2 in the far right pocket.

Sorting Board

No. of statements for each position

Most like Colgate	2	3	5	5	6	6	6	5	3	2	Least like Colgate
	__	__	__	__	__	__	__	__	__	__	

The respondent repeats the process for the other brands being rated. One complexity of the Q-sort technique arises from the initial selection of 48 statements. These are chosen from a larger pool of statements, much as in the Thurstone method.

A variation of the method lies in having respondents sort out a large number of *brands or products* into 11 piles, rather than sorting out 48 statements regarding one brand. (Few products, however, have 48 brands which can be used.)

Another usage is in pre-testing ad headlines or sales propositions or themes. Forty-eight different ad headlines for a particular brand can be written, for example, and these sorted out into 11 piles along various dimensions: persuasive to non-persuasive, interesting to dull, etc. This usage more properly belongs in the pre-testing chapter, however.

Hierarchy scales. A hierarchy of attitudes, or product preference, has already been defined as a "ladder" of successively more favorable attitudes. The highest point on the hierarchy with which you agree implies that you would also agree with or endorse all lower points on the ladder, and reject all higher points on the ladder.

One such hierarchy, a "ladder of adoption of new products" is: awareness, interest, evaluation, trial, adoption. This theory, based on sociological research, postulates that people appear to go through a series of distinguishable stages in adopting a *new* idea or product--awareness, interest, evaluation, trial and adoption. *Awareness* is the knowledge that a new idea, product or practice exists; *interest* is the active seeking of more information to determine its possible usefulness; *evaluation* is the weighing and sifting of that information, either mentally or in discussions with others; *trial* is the tentative trying of the product; *adoption* is the full-scale continuing integration of the product into attitudinal and buying processes.

Psychologists have postulated another three-stage hierarchy of human behavior--cognition, affection, and conation. Translated from the technical jargon, the theory is that people pass through successive stages of *awareness* and *factual knowledge* (cognition), favorable *attitudes* and feelings (affection), strong motivation or *desire for action* (conation), possibly eventuating in the *action* itself--in the case of advertising, purchase of the product.

This illustrates another assumption of hierarchy scales or models--that a person must pass through each stage of the process in order, without skipping any stages.

A similar "ladder of product preference" developed by advertising practitioners is the *AIDA* hierarchy: attention, interest, desire and action.

One large automobile manufacturer has conducted continuing attitude surveys to determine, among other things, consumer position on a product preference ladder. The marketing process for any particular make of car is visualized as one of moving people up the ladder from *unawareness*, to *awareness*, to "knowledge that the car-make is in a particular product class," to "consideration of the make as among those makes you might buy," to "first choice."

Another major automobile manufacturer has used a similar but simpler "buying probability" hierarchy based on consumer surveys. For the highest to lowest rung, it reads:

Would probably buy
Would strongly consider
Might consider
Would definitely not consider
Aware that it exists

The hierarchy scales have one principal distinction from the other kinds of scales discussed. A person's position on the scale is not determined by a single question in which he checks one or more responses. Rather, a series of questions is asked and a person's place in the hierarchy is determined by cross-tabulation of all the responses. This can be illustrated in the case of the "buying probability" scale by the following survey questions:

1. When you think of new automobiles, which make comes to mind first?
2. What other makes come to your mind?
3. If you were going to buy a new car today, which make would you probably buy?
4. What other makes would you want to shop before making a final decision?
5. Which makes, if any, would you definitely not consider buying under any circumstances?

In summarizing the results of the above questions to establish the distribution of consumers on the buying probability hierarchy, the data would be compiled in this way for each car make:

Level of Hierarchy	Those persons naming Car X
Awareness	Q. 1 or 2
Definitely not consider	Q. 1 or 2 and Q. 5
Might consider	Q. 1 or 2 and NOT Q. 5
Strongly consider	Q. 1 or 2 and Q. 3 or 4
Probably buy	Q. 1 or 2 and Q. 3

Here's how the responses to such a survey among persons identified as prospects, might be summarized in the case of three fictious competitive cars--the Constellation, the Chevalier, and the Mayflower.

% of all car prospects who...	Make of Car		
	Const.	Chevalier	Mayflower
(a.) Are aware	85%	90%	75%
(b.) Would definitely not consider	45	20	40
(c.) Might consider	40	70	35
(d.) Would strongly consider	38	50	25
(e.) Would probably buy	25	35	15

Note that the proportion who fall into categories b. and c. add up to the number who are aware. Another category, not shown, is the "unaware" group--15%, 10%, and 25% respectively for the three makes. The total for the three makes in category e. adds up to 75%--obviously 25% of the prospects have named other makes.

Hierarchy scales are not, strictly speaking, attitude scales; they can incorporate a number of dependent variables effects ranging all the way from awareness to behavior reports, with the strict definition of a favorable attitude being only one "effect" that might be desirable.

In concluding the discussion on hierarchy scales, it should be noted that the "hierarchy" theory has not been validated--that is, no one has definitely established that everyone passes through the same stages up the ladder of product preference, nor that these are the crucial stages even if there is a hierarchy. Like much other valuable marketing data, most advertisers would try to keep it secret even if they had developed a valid scale. However, the concept is exceedingly useful in analyzing the marketing and advertising process and should not be discounted simply because it has not been thoroughly validated. In fact, very few areas of human knowledge or measurement have been thoroughly proven to everyone's satisfaction.

Hierarchy scales are the only tools discussed here which are not suitable for image measurement. All the others--monadic scales, relative preference scales and complex scales--are equally applicable to both attitude and image measurement.

Controlled Observation

Sometimes it is possible to set up a situation in which people are watched without their knowledge while they view television, listen to radio, read newspaper and magazines, or perform other activities. Their reactions may be observed through a one-way mirror, which appears to the respondent to be a mirror but is transparent from the other side. Or hidden cameras can record their reactions. Or someone in the same location can surreptitiously observe and record subject behavior while ostensibly doing something else. Or hidden microphones can record comments.

When a live observer is used, his observations may range from impressionistic description of behavior to highly quantified ratings and precoded categories. Filmed and recorded observations have the advantage of being all-inclusive and permanently available for analysis by several researchers.

Observation, if properly controlled, has the advantage of being an "unobtrusive measure," not alerting the respondent that he is being "tested". This, of course, results in data that are uncontaminated by artificial testing conditions or verbal camouflage.

Distinguishing characteristics. The research environment, the data collection method, and the response method are the principal distinguishing characteristics of this plan.

Research environment: Stimulus exposure, staged; response environment, surreptitious

Data collection method: personal recording and mechanical recording of selected responses.

Response method: Conversation, facial expressions, gestures, bodily movement, focus of attention; spontaneous and unstructured.

Uses of controlled observation. One can, of course, observe responses to advertising in any situation where it naturally occurs. For most useful results, however, the stimulus exposure should be "staged" so that the respondent is exposed to particular media or advertising of interest to the researcher but in an environment where the subject doesn't suspect that he is being deliberately exposed and observed.

One example of such research is the classic "peek-a-boo" study conducted by Curtis Publishing Company many years ago. Adults were invited to take part in an "eye test." After appearing at the designated address, each respondent was required to wait his turn for the "test" in a specially designed waiting room. The waiting room contained a comfortable chair, beside which was a magazine rack and reading lamp. Behind the chair, unknown to the subject, was a one-way mirror; an observer watched the subject without his knowledge. As the subject killed time by looking through the magazines, the observer was able to record which ads and magazine articles captured attention, the amount of time that was spent with each, and other reactions. In such a situation, certain effects of specially planted ads can be determined.

The "candid camera" technique is another application. By staging the kinds of ads or posters that are exposed to unknowing subjects in a fixed location, a hidden camera can make a record of their focus of attention, conversation, and other reactions. By planting a confederate on the scene to ask seemingly casual questions, more specific verbal responses can be obtained.

A device called the "Dynascope" has been used to measure attention to television programs and commercials. This consists simply of a camera mounted on top of a television set, to record attention to programs and commercials. This is not exactly surreptitious observation, but the reasoning is that after a while the viewers will tend to forget the presence of the camera and react naturally.

Another advertising researcher used graduate students as "spies" in their own homes to observe and record responses of family members to television programs and commercials. While ostensibly reading books and magazines, the students were actually watching and recording on predesigned forms the reactions of their unsuspecting family. This technique has also been used to obtain behavioral reaction to newspapers and magazines.

Another application of controlled observation is reported in a study and analysis of public conversations. The investigator simply stationed himself in various public gathering places and recorded the spontaneous comments of persons nearby. The variety of topics covered by the conversationalists, however, rarely included advertising and would hardly ever include particular advertising of interest to the researcher; one way to circumvent this problem is to plant particular advertising stimuli in hopes that the people nearby will notice and comment upon it.

Controlled observation is applicable only for certain specific problems of advertising research. It does have advantages over many other plans, notably in that subjects can be exposed and measured without their knowledge.

Content Analysis

The end product in the creation of advertising is the advertisement itself. These physical products take the form of printed ads for newspapers, magazines, outdoor or direct mail media; for the broadcast media of radio and television, the physical product may be permanently recorded on film, video-tape or sound recordings. This physical manifestation of the message--the resultant of all the creative effort, business planning and research pre-testing--can be studied and analyzed for various purposes.

The method for such study is known as content analysis. Content analysis is a research plan for the objective, systematic and quantitative description of the manifest content of advertising or other forms of communication. The key words in this definition are objective, systematic, quantitative and manifest. These words distinguish scientific content analysis from the ordinary informal analysis of everyday life in watching television, listening to radio or reading newspapers and magazines.

Systematic means that the selection of advertising for analysis must be based on a formal, predetermined, unbiased plan. This insures that it will not be simply an argumentative, biased collection of data to prove a point. Quantitative means that the results of the analysis are usually expressed numerically in some way--percentages, ratios or more complex statistical summaries. Results are more meaningful when one can say "The theme of 'speed' was used in 30 per cent of Ford newspaper ads and in 20 per cent of Chevrolet ads" rather than the vague statement "Ford appears to use the speed theme more often than Chevrolet." Objective means that content categories or codes for analysis have been defined beforehand, and that these are defined so precisely that different coders would classify the content in the same way. Manifest content refers to those ad elements that can be actually seen and counted rather than reading "between the lines" for hidden meanings--the denotative content rather than any possible connotations of what has been recorded.

Distinguishing features. From this general discussion, it will be gathered that content analysis differs in one crucial respect from other advertising research plans previously discussed; it is not human behavioral research, that is, the object under direct study is advertising content itself, not the response of people to advertising.

Data collection method: observation of content
Design method: after-only
Dependent variable: any aspect of pictorial or verbal or organizational presentation
Sampling: complete enumeration of population of ads, or random sampling from prescribed vehicles and vehicle units.

Uses of content analysis. The analysis of published or broadcast advertising has several research applications.

There are certain general principles of communication, learning, persuasion, visual perception, readability and motivation which have been reasonably well established through basic psychological research. A listing of these principles, together with a statement of advertising objectives, can be useful as a guide to both the creation and evaluation of advertising. In pre-testing, the actual content of rough or finished ads can be checked against these principles; where the ads don't fit the prescription, they can be modified.

Content analysis has several procedural advantages over other forms of advertising research. The basic research material, in the form of printed ads or recorded commercials, is readily available, permanent and doesn't require human subjects. These factors add up to savings in time and money over other research methods. It cannot serve as a substitute for behavioral research, but it is a useful adjunct sometimes when research with human subjects is impractical.

Readability analysis is one familiar form of content analysis. The basic factors involved are word length and sentence length; it has been demonstrated that verbal messages using short words and short sentences are read more quickly, comprehended better and more likely to be read completely than the same message using long sentences and long words. Rudolf Flesch has developed a formula for quickly analyzing and giving a numerical readability score to written materials.

Various other formulas dealing with other aspects of readability--for example, the Gunning "Fog Index," Gillie's Abstraction Index and the Dale-Schall Index--have also been developed.

Television scripts can also benefit from content analysis. The Dupont company's advertising research department developed a coding system based on psychological principles; content analyses were used to make predictions of effectiveness and changes were made if a script seemed to be lacking.

Some basic research and experimentation in print ad layout and design has been conducted by the Advertising Research Foundation and the HRB-Singer Company. This led to the formulation of several layout principles for high "attention value" against which ads can be compared. The amount of white space, color usage and other factors are important ingredients which can be checked by content analysis.

Another usage of content analysis is in the evaluation of competitive advertising. If there is reason to feel that a competitor's newspaper advertising has been unusually successful, for example, content analysis of themes, persuasion strategies, choice of words, or use of illustrations and color can produce some tentative conclusions on why the advertising was successful.

Or an advertiser might analyze his own brand's advertising over a certain period to determine if pre-determined advertising themes or other content goals have been met.

Another usage is to compare high-readership ads and low-readership ads. If quantitative differences in content are found, the researchers may conclude--subject to further test--that the observed differences might be a cause of the difference in effectiveness. (These "possible causes" are only hypotheses until verified by experimentation.)

From the above discussion, it is seen that content analysis can be used in both pre-testing and post-testing messages. It is no substitute for behavioral research, since it provides no direct evidence of advertising's effects on people. But, combined with behavioral research, it can help provide understanding of why ads have performed well or poorly. Additional discussion of content analysis is found in the message research chapters.

IV. MEDIA RESEARCH IN GENERAL

An advertiser who allocates a certain amount of money for advertising is, quite naturally, interested in getting the most for his money. Part of the advertising dollar goes into development of messages, another part into the media which carry the messages.

Determining the most efficient media is the primary function of media research--to measure and compare different media and different vehicles in terms of vehicle distribution, vehicle exposure and advertising exposure. Media research must provide quantitative data from which objective decisions can be made. No advertising budget ever devised, even in the largest corporation, would permit use of more than a fraction of all the advertising time and space available. Out of the innumerable combinations of media vehicles available, something more than guesswork is needed to put together that particular "media mix" or "media schedule" which is most efficient in reaching a product's target audience.

What is media, or media-vehicle, efficiency? It is a function of both the number of product-prospects reached and the cost of reaching them. A highly efficient vehicle is one which reaches a great number of prospects at low cost. The usual measure of media efficiency is "cost-per-thoudand (CPM)." This can be further specified in terms of (a) vehicle distribution as cost-per-thousand circulation units obtained, (b) vehicle exposure as cost-per-thousand prospects exposed to the vehicle, and (c) advertisement exposure as cost-per-thousand prospects exposed to the vehicle space carrying the advertisement.

Questions Involved in Choosing Media

Here are some of the questions that occur to an advertiser at this point.

Which of the media--television, radio, newspapers, outdoor, direct mail, skywriting, matchbook covers--are best for this particular product? Should the amount of advertising be spread equally among the chosen media, or should some get more than others? How much more should the favored ones get? Should the advertising appear simultaneously in the chosen media, or should the usage be alternated among them at different times?

Assuming the amount for a given medium is decided upon, how should it be allotted among the many vehicles in that medium? In any particular vehicle, should the advertising be spread out equally over time, or bunched in some time periods and eliminated in others? Is it better to use many small ads, or a few large "blockbusters"?

All these are questions that must be resolved by media research.

Typical Media Research Studies

A good way to get the feel of media research is to look at the titles of some of the research studies already performed recently. Most of the studies reported below appeared in the Journal of Advertising Research, an official publication of the Advertising Research Foundation.

Media research studies can deal with a single vehicle or a comparison of several vehicles, with a single medium or a comparison of several media, with vehicle "mix" studies or media mix studies, with media and vehicle "image" studies, and with reviews and discussions of advertising strategy and research methods. The unfamiliar terms should become known to you through studying this and the succeeding sections.

"Single vehicle" studies are those which report on the effectiveness of one vehicle. Agostini's article "Analysis of Magazine Accumulative Audience" tells how the accumulative audience of several successive issues of a magazine can be determined from knowing the audience size for two issues. Another Agostini article "The Case for Direct Questions on Reading Habits" was a comparison of two methods for determining how often a person reads a particular magazine. Bogart's article "The Impact of Blank Space: An Experiment in Advertising Readership" tells how the readership of a newspaper advertisement is affected by other material appearing on the same page, or the influence of "context." Engleman's article "An Empirical Formula for Audience Accumulation" shows how the accumulated audience for 13 issues of a magazine can be estimated with a simple formula.

It should be emphasized that most of the studies reported above, and those that follow, are based on actual empirical research and are not merely speculations or armchair discussions; the essence of effective decision making lies in the use of scientific evidence, and not in unsupported opinions.

"Multiple vehicle" studies report on comparisons of several vehicles with each other, and on the effectiveness of various combinations or "mixes" of vehicles. Agostini's article "How to Estimate Unduplicated Audiences" presents a formula for estimating the combined audience of a magazine mix, from information about the individual magazines' audiences. Day's article "Linear Programming in Media Selection" discusses the advantages and disadvantages of using a computer mathematical technique, linear programming, in picking from a long list of vehicles those few that will represent the best mix for reaching a desired audience. Bower's article "Net Audiences of U.S. and Canadian Magazines" is an additional study to verify Agostini's formula for estimating the size of unduplicated audiences for a particular combination of magazines. Caffyn and Sagovsky's article "Net Audiences of British Newspapers" is another test of the same formula applied to newspapers, illustrating that both newspapers and magazines can make use of the same research methods. (As will be illustrated later, however, the print media and the broadcast media must use different research methods and measurements and this results in some difficulty in making a direct comparison between them.) Garfinkle's article "A Marketing Approach to Media Selection" presents a method for comparing various vehicles in terms of the number of present customers and prospective customers reached. Kuhn's article "Net Audiences of German Magazines" points up the importance of knowing how to determine the unduplicated or net audience of a vehicle mix. Marc's article "Net Audiences of French Business Papers" deals with vehicles for specialized audiences rather than mass audiences. Corlett's article "The IPA National Readership Study" shows how the audiences for 80 British newspapers were determined in a single-interview survey. Dudek's article "Relations Among Television Rating Indices" tells how the value of different television programs is dependent not only on the size of their audiences but also on how much they like the program. Massy's article "Discriminant Analysis of Audience Characteristics" tells of a statistical technique for comparing vehicles, showing that some vehicle mixes reach the same persons while others reach quite different segments of the public.

There are fewer studies of a whole medium, or several media, than of vehicles, for the simple reason that it's difficult to make a single generalization that will apply to the many diverse vehicles within any single medium. Research on media vehicles can thus be more precise than on whole media, despite the need to know about the advantages and disadvantages of advertising in the various media. Ziff's article "How to Locate Purchasers of a Product" points out that to select the best media for advertising to current customers, you have to know who those customers are and what media they use. Buchanan's article "How Interest in the Product Affects Recall: Print Ads vs. Commericals" compared print and broadcast media advertising effectiveness in terms of teaching product facts. Landis' article "Exposure Probabilities as Measures of Media Audiences" reports on a measure of exposure designed to solve a knotty problem in media research, the direct comparison of TV program and magazine audience size.

"Media image" studies are aimed at determining how one's attitude toward, or image of, a vehicle or medium can enhance the perception and communication value of a given message. For example, Magazine A and Magazine B may deliver the same number of advertising exposures for the same message among the same people, but if Magazine A has a better reputation for honesty, the ad space may automatically deliver greater believability for any message placed there. The attitudes toward or images of media are sometimes called "qualitative" variables to indicate that they have no precise numerical value. Weilbacher's article "The Qualitative Values of Advertising Media" discusses several kinds of qualitative values of a vehicle--audience characteristics, editorial personality, contribution to advertising effectiveness--and evaluates those qualities in terms of research techniques, media comparisons, vehicle comparisons, and media selection. Meissner's article "Sales and Advertising of Lettuce" compares the attitudes toward different media and the relationship of those attitudes with sales. Kotler's article "Toward an Explicit Model for Media Selection" demonstrates a way to assign numerical values to qualitative image information, so that it can be figured in with quantitative data in evaluating vehicles or media.

Among the most valuable articles on media research are those that deal with research methodology. Such studies determine the validity and practicality of different research methods in obtaining accurate data. Just as a watch must be an accurate instrument in measuring time, so an advertising research tool must be an accurate instrument for its measurement task. Cahalan's article "Measuring Newspaper Readership by Telephone" compares telephone interviews with personal interviews as data-gathering methods for determining magazine vehicle audience size. Stock's article "A Comparison of Eight Audience Estimates" compares the audience figures obtained with eight different kinds of questions relating to magazine readership. Copland's article "Some Fundamentals of Poster Audience Measurement" deals with the peculiar problems involved in determining outdoor advertising exposure, and compares several different methods of measurement. Kirsch and Banks' article "Program Types Defined by Factor Analysis" describes a statistical method for determining TV program types and the audiences for those types. Henry's article "Belson's Studies in Readership" shows how the questionnaires and interviewing methods used in determining audiences of print media can create large errors in the final figures. Allen's article "Photographing the TV Audience" illustrates one method for determining the difference between circulation and audience of a TV commercial, this study showing that half the time no one was watching even though the set was turned on. Zangwill's article "Media Selection by Decision Programming" is a comparison of two computer methods in determining vehicle mix audiences.

After this broad overview of the kinds of studies that go under the name of media research, it's time to turn to the research methods used. The print media and the broadcast media will be considered separately, since somewhat different methods must often be used for each.

selection. Meissner's article "Sales and Advertising of Lettuce" compares the attitudes toward different media and the relationship of those attitudes with sales. Kotler's article "Toward an Explicit Model for Media Selection" demonstrates a way to assign numerical values to qualitative image information, so that it can be figured in with quantitative data in evaluating vehicles of media.

Among the most valuable articles on media research are those that deal with research methodology. Such studies determine the validity and practicality of different research methods in obtaining accurate data. Just as a watch must be an accurate instrument in measuring time, so an advertising research tool must be an accurate instrument for its measurement task. Cahalan's article "Measuring Newspaper Readership by Telephone" compares telephone interviews with personal interviews as data-gathering methods for determining magazine vehicle audience size. Stock's article "A Comparison of Eight Audience Estimates" compares the audience figures obtained with eight different kinds of questions relating to magazine readership. Copland's article "Some Fundamentals of Poster Audience Measurement" deals with the peculiar problems involved in determining outdoor advertising exposure, and compares several different methods of measurement. Kirsch and Banks' article "Program Types Defined by Factor Analysis" describes a statistical method for determining TV program types and the audiences for those types. Henry's article "Belson's Studies in Readership" shows how the questionnaires and interviewing methods used in determining audiences of print media can create large errors in the final figures. Allen's article "Photographing the TV Audience" illustrates one method for determining the difference between circulation and audience of

Measuring Distribution of Print Media Vehicles

No original research is needed to determine distribution, defined as the number of copies circulated. The Audit Bureau of Circulations (ABC) provides an actual count of the number of copies of a newspaper or magazine delivered under specific conditions. ABC was founded to provide assurance to advertisers that the circulation reported by a publication is accurate. Since the cost of print advertising is based on the number of copies in which an advertisement appears, this basic figure is important but doesn't go too far toward measuring vehicle effectiveness. ABC follows a very rigorous bookkeeping and verification procedure to insure that a newspaper or magazine delivers the number of copies contracted for. The advertising researcher can rest assured that circulation figures are accurate as reported by the ABC. No more need be said here about the methods used. Some fictitious figures for magazines and newspapers are given below:

	Magazine A	Magazine B
Cost, 1-page non-color ad	$35,000	$50,000
Circulation (No. of copies)	7,000,000	10,000,000
Circulation efficiency (Cost per 1000 copies)	$5.00	$5.00

The example above illustrates that, while the total cost of an advertisement space may vary considerably from vehicle to vehicle, it is generally proportional to circulation within any given medium, and the circulation cost-per-thousand will not vary too much except under certain conditions of specialized audiences, preferred treatment, etc.

Measuring Issue Audiences of Print Media Vehicles

The research methods for measuring audience exposure are about the same for magazines and for newspapers. Consequently, this section will be primarily a discussion of magazine audience research, with the understanding that the discussion is really about both kinds of the major print media. The reason for this approach is two-fold: more research has been published on magazines than newspapers, and it is a good mental exercise for the beginning media researcher to take a given magazine example and think it through for a newspaper example.

The principle variables to be measured for determining audience exposure to a magazine or newspaper are:

Audience size: the number of persons reading the average issue, or total issue audience

Audience composition: the *kinds* of persons in the issue audience, in terms of demographic characteristics such as age, sex, education, income, etc.; product buying and usage characteristics such as heavy users, average users, and non-users; other pertinent audience characteristics

Amount of audience exposure: the amount of issue exposure by individuals in the audience--number of times the issue is picked up, the number of days it is read, the number of items read and so on; also, readership of different specific issues of a single magazine.

How is such information obtained? The *sample survey* is the principal method. Personal interview surveys, telephone surveys, and mail surveys have all been used for audience research. Under some circumstances one method will be better than another, under other circumstances the mail, telephone and personal methods are equally good. Much methodological research has been devoted to determining which method is best under what circumstances.

Planning and executing a media research survey involves a great deal of painstaking attention to the numerous details of survey procedure described earlier. The most crucial concerns in audience surveys are those of sampling and question writing--the ultimate validity of the data rest heavily on these two aspects of the procedure.

Sampling. The first step in sampling for media research is to define the crucial population. The crucial population is dependent on the purpose of the study. In a single-vehicle study (one magazine or one newspaper), the crucial population is "all persons in the circulation area who have read at least one issue" in a specified time period; if a complete list of magazine buyer's names is available, it may be a relatively simple matter to draw a random sample of persons from that list.

On the other hand, in a multiple-vehicle study in which the purpose is to determine the individual and overlapping audiences of a number of vehicles, the procedure is more complicated; the crucial population is defined as "all persons reading at least one issue of one magazine on the list" in which case it is more practical to select a random sample of the entire U.S. population, and use screening questions to weed out those who don't qualify as a reader. In the case of specialized magazines, such as those intended for physicians, a complete listing of this special population may be available through a professional organization or a mail-list firm. In such a case, either single-vehicle or multiple-vehicle studies can utilize simple random sampling from the list.

Again, in the case of a particular product whose buyers are known, the advertiser may define his crucial population as "all owners or users of the product"; if a listing of that population is available, a simple random sample of those persons can be drawn from the list more easily than from a sample of the entire population. In the case of a diaper manufacturer, for example, a listing of his crucial population of recent mothers may be obtainable; similarly, a truck manufacturer can obtain a listing of truck owners for surveying the media habits of that crucial population.

Most major media studies, however, are multiple-vehicle studies of the entire U.S. adult population, in which case listings of the crucial population are either unobtainable or inconvenient. Such studies are often sponsored and paid for by several advertisers, with the actual research carried out by a commercial research firm. From these "syndicated" studies, separate tabulations of the results are then tailor-made for individual advertisers or media.

For illustrative purposes, let's focus here on a personal interview sample survey of the entire U.S. adult population, in which data on several magazines and on several products, are gathered. Assuming that a completely random and representative sample of this population has been drawn, the validity of the questions becomes the crucial concern.

Question writing. An operational definition of "a magazine reader" must be arrived at before the proper questions can be phrased. Is a "reader" someone who (a) has seen the cover or (b) has riffled through the pages or (c) has read at least one item, or (d) someone who usually reads it, or (e) what? Defining readership is a game any number can play, but the usual definition of a reader is someone who has at least looked into an issue. From this, it follows that magazine audience size is the number of people who look into the average issue. Data from several issues is usually necessary to determine the average issue.

Three general types of questions for determining reading behavior are unaided recall, aided recall, and recognition.

"Unaided recall" questions on issue reading are those where no clues are given, and the respondent must rack his memory for the answer. An example of an unaided questioning sequence would be:

(1) Do you ever read any magazines?
(2) Which magazines have you read during the last year?
(3) IF MAGAZINE A IS MENTIONED: Did you happen to see the March 21st issue of "A"?
(4) IF YES: Did you happen to look into it?
(5) Can you remember what was on the cover?
(6) Can you remember anything you saw in that issue?

Unaided questions 1-4 rely on the respondent's memory, willingness to think about a relatively insignificant event, desire to please the interviewer and many other factors which may lead to an answer which is consciously or unconsciously distorted. Questions 5 and 6 attempt to get some factual verification of the readership claim, but these also leave doubts as to reliability. Can *you* remember what was on the cover of the third-from-last issue of *Time* magazine, or something specific inside that issue?

"Aided recall" questions on issue reading give the respondent more information on the magazines in question and produce different results. An example of an aided questioning sequence would be:

(1) I'm going to read off the names of several magazines. As I name each one, please tell me if you have seen any issue of it during the last year. (INTERVIEWER READS LIST OF MAGAZINE NAMES TO RESPONDENT)
(2) (FOR EACH MAGAZINE SEEN BY RESPONDENT, INTERVIEWER ASKS:) Did you happen to see the last issue of "A"? It had a picture of ____________ on the cover?
(3) Can you remember anything you saw or read in the issue?

Question 1 provides the respondent with specific magazine names and produces more readership claims than the unaided sequence. Question 2 aids the respondent in visualizing a particular issue, and it would produce a larger number of claims than the unaided sequence. Question 3 attempts to verify reading of a particular issue by getting the respondent to recall (unaided) something he saw in the issue; such a question produces underestimates of audience size, since it has been amply demonstrated that many known readers of an issue can't recall anything in particular from it--or can't be sure that something they recall reading came from a particular issue.

"Recognition" questions provide the respondent with even more information on magazine issues, by actually showing him the logotype, or the cover or perhaps some of the contents of magazine issues. (The logotype is the standard format showing the magazine's name on every cover.) A recognition questioning sequence might run something like this:

(1) Here are the logotypes of several magazines. As I show you each one, please tell me if you happen to have seen any issue of it during the last year.
(2) (FOR EACH MAGAZINE MENTIONED BY RESPONDENT, INTERVIEWER PRODUCES A PARTICULAR ISSUE OR ISSUES AND SAYS:) Here is an issue of "A" that you may not have happened to read yet. Do you remember whether or not you have looked into this particular issue? Let's leaf through the first few articles to see if you recognize anything.
(3) (AFTER LEAFING THROUGH PART OR ALL OF THE ISSUE, INTERVIEWER SAYS:) Just for the record, now that we have been through this issue, would you say you definitely happened to read it before, or didn't read it, or aren't you sure?

Only those persons who say definitely they remember reading the issue are counted in its audience. An additional technique for eliminating persons who claim issue reading falsely is the "confusion control" technique. Several dummy items that didn't actually appear in the issue are inserted in the interviewer's copy; if the respondent mistakenly claims to have seen the dummy items, he may be eliminated from the subsequent tabulations or his responses might be given reduced weight in the overall tabulations, to allow for his "confusion" or exaggerated claims.

The "recognition" technique for determining issue audiences is considered by many to be more accurate than either the unaided recall or aided recall techniques. It has been compared with other data from known issue readers and found to be a practical method for comparing the audiences of several vehicles. However, much methodological research is still underway to produce more precise and accurate audience figures.

Audience efficiency illustration. The audience for several issues is averaged to produce figures on "average issue audience" to provide a more stable figure. With data on average issue circulation and average issue audience, it's possible to make some additional judgments on vehicle value for the advertiser. Here's the previous example using circulation figures, with audience figures added, and an audience efficiency computation. "Efficiency," as in the previous example, is the cost-per-thousand, abbreviated to CPM, which makes it possible to compare on the same basis different vehicles with varying circulation or audience or other measures.

	Magazine A	Magazine B
Cost, 1-page non-color ad	$35,000	$50,000
Circulation (No. of copies)	7,000,000	10,000,000
Circulation efficiency (CPM copies)	$5.00	$5.00
Readers per copy (average)	3.36	2.50
Issue audience (No. of readers)	23,500,000	25,000,000
Audience efficiency (CPM readers)	$1.49	$2.00

The example above shows why audience figures are better for comparing magazines than are circulation figures. On the basis of circulation, one would assume that Magazine A and Magazine B are equally efficient for the advertiser. But because Magazine A has more readers per copy, it produces readers at a reduced cost-per-thousand--i.e., Magazine A is more efficient in terms that are meaningful to the advertiser.

Prospect audience efficiency illustration. Even more important to the advertiser, however, is the number of prospects for his product who are in the issue audience; money invested in advertising to people who are not current or future prospects for the product is wasted.

The sample survey above, in which the issue readership data were obtained, also obtained other data on demographic characteristics (age, sex, occupation, income, etc.) and on product buying and usage. This additional information can be used by advertisers to get an even better estimate of the efficiency of particular vehicles for particular products.

This is illustrated below by looking at the prospect audience for three advertisers: a soap powder, an airline, and a teenage pimple remedy.

In the case of the soap powder product, the market or prospects are defined as "female adult household heads." For the airline, prospects are defined as "adults between the ages of 25 and 60 who have ridden a commercial airline in the last two years." For the pimple remedy, prospects are defined as "13-19 year olds with allowances or incomes over $3 per week." (The prospect definitions were arrived at by previous surveys which correlated product buying with demographic characteristics.)

	Magazine A	Magazine B
Ad cost	$35,000	$50,000
Circulation	7,000,000	10,000,000
Circulation efficiency (CPM copies)	$5.00	$5.00
Total audience (average issue)	23,500,000	25,000,000
Audience efficiency (CPM readers)	$1.49	$2.00
SOAP POWDER PROSPECTS		
% of total audience	25.5%	34.3%
No. of prospects	6,000,000	8,580,000
Prospects per copy	.85	.86
Prospect audience efficiency (CPM prospects)	$5.83	$5.34
AIRLINE PROSPECTS		
% of total audience	11.9%	31.0%
No. of prospects	2,800,000	8,000,000
Prospects per copy	.40	.80
Prospect audience efficiency (CPM prospects)	$12.50	$6.25
PIMPLE REMEDY PROSPECTS		
% of total audience	38.7%	20.0%
No. of prospects	9,100,000	5,000,000
Prospects per copy	1.3	0.5
Prospect audience efficiency (CPM prospects)	$3.85	$10.00

The illustration above shows that Magazines A and B are equally efficient for reaching the soap powder market. Magazine B is considerably more efficient for reaching the airline market, $6.25 CPM vs. $12.50. Magazine A is considerably more efficient for reaching the pimply-faced market, $3.85 CPM vs. $10.00.

The major points are that the prospect-audience is a more useful group for selecting media vehicles than the total audience, and that each vehicle must be evaluated in terms of its efficiency for a particular product. There is no such thing as "the best advertising vehicle" but only "best advertising vehicles for particular products."

In addition to "total issue audience" which tells how many people have the opportunity for exposure to a single advertisement insertion in a single issue, the advertiser is also interested in the "accumulative audience" for several issues of the same publication, and the "unduplicated audience" for several different publications.

Accumulative audience. Different issues of the same publication always exhibit some audience turnover. That is, of the people who read the June issue of a monthly publication, some will not read the July issue; and some persons who didn't read the June issue will read the July issue. The "audience turnover" varies from one magazine to another; this means that two magazines with an equal average issue audience may have different accumulative audiences over a period of time. Accumulative audience may be defined as the number of different people reached by two or more issues.

Magazines A and C, both monthlies, have identical audience size for each month in the table below, and they have an identical average issue audience, but their accumulative audiences are quite different, 34 million for A and 29 million for C.

	Magazine A	Magazine C
June audience	24 million	24 million
July audience	23 million	23 million
August audience	24 million	24 million
September audience	23 million	23 million
Total audience	94 million	94 million
Average Issue	23,500,000	23,500,000
Accumulative audience	34,000,000	29,000,000

How could such a great difference in accumulative audience occur? Because Magazine C has less audience "turnover"--that is, fewer people drop out of and come into each suceeding issue. The accumulative audience is the total number of people who read any one or more of the four issues.

Here's how the 34 million accumulative audience was obtained for Magazine A.

Magazine A Audience (in millions)

Issue	Issue audience	Old audience	Accumulative drop-outs	New audience	Accumulative audience
June	24	--	--	24	24
July	23	20	4	3	27
August	24	21	6	3	30
September	23	19	11	4	34

"Old audience" is defined as the number of this issue readers who have read at least one previous issue in the series.

"Accumulative dropouts" is defined as the number of this issue non-readers who have read at least one previous issue in the series.

"New audience" is defined as the number of readers of this issue who have not read any previous issue in the series.

An issue audience consists of "old readers" plus "new readers."

The accumulative audience in any series consists of the "old audience" plus the "accumulative dropouts" plus the "new audience."

For June (the first issue of the series), there is no old audience and no accumulative dropouts, consequently the 24 million new readers = issue audience = accumulative audience.

For July, there are 20 million old readers from June, 4 million dropouts from June, and 3 million new readers. The July issue audience is thus 23 million, The June-July accumulative audience is 20 million old plus 4 million dropouts plus 3 million new, for a total of 27 million.

For August, there are 21 million old readers from June-or-July, four million June dropouts and two million July dropouts (six million accumulative dropouts), and three million new readers. The August issue audience is thus 24 million. The June-July-August accumulative audience is 21 million old plus 6 million accumulative dropouts plus 3 million new, for a total of 30 million.

The reader may test his understanding of this procedure by going through the same process in arriving at the June through September accumulative audience of 34 million.

In the table below, the audience information for Magazine A is compared directly with that from Magazine C to explain how the former has a much larger accumulative audience. That is to say, A reaches more different people.

Audience in Millions

	Magazine A				Magazine C			
Issue	Old aud.	Drop outs	New aud.	Accum. aud.	Old aud.	Drop outs	New aud.	Accum. aud.
June	--	--	24	24	--	--	24	24
July	20	4	3	27	22	2	1	25
August	21	6	3	30	22	3	2	27
Sept.	19	11	4	34	21	6	2	29

It's worth taking the time to master the simple arithmetic involved in this example in order to understand the concept of accumulative audience.

This example will serve to illustrate two other concepts in media research, "reach" and "frequency."

Reach. Reach is defined as the number of people reached, and it is identical to audience. Thus, the reach of the individual issues, or of the average issue, is identical for Magazines A and C. However, Magazine A has greater reach than C if we're referring to the total number of people reached with a series of four issues.

Frequency. Frequency refers to the number of times each person is reached. Thus, while Magazine A has greater accumulative reach, Magazine C achieves greater frequency. This is illustrated in the table below.

	June-September Issues	
	Magazine A	Magazine B
Average issue audience	23,500,000	23,500,000
Total audience	94,000,000	94,000,000
Accumulative audience	34,000,000	29,000,000
Frequency of reach	2.76	3.24

This table shows that the average reader in the Magazine A accumulative audience read 2.76 of the four issues, while the average reader for Magazine C reads 3.24 issues--that is, his frequency of reading was greater.

This raises the question as to which is more desirable to the advertiser, reach or frequency? To use a simplified example, is it better to reach 100 million persons with one exposure each, or to reach 50 million persons with two exposures each? In each case, the total number of audience exposures is the same and assume the cost is the same. This in turn leads to the question, how many exposures are necessary for advertising to have an effect on a prospect? If one exposure is adequate to produce a sale, then reach is the desired measurement; however, if two exposures are required to produce a sale, then frequency becomes a crucial measure. The truth is, no one really knows just how many exposures are required to produce an effect; it probably varies for different kinds of products and for different kinds of persons. This is one illustration of the need for basic research which would discover the buying probabilities created by a varying number of exposures to product advertising.

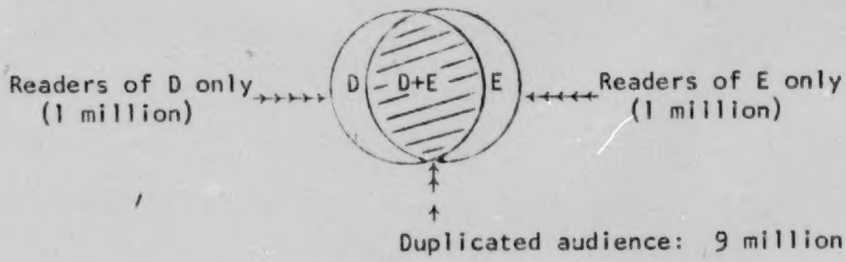
One pioneering study in this area was conducted by the Saturday Evening Post to determine if the second exposure to an ad was less valuable or more valuable than the first exposure. A very complicated and ingenious controlled experiment was conducted which revealed that the second exposure by a person to a particular advertisement was if anything even more effective than the first. However, a great deal more of such research is needed to be able to assign a precise value to exposure number 1, number 2, number 3 and so on in order to determine the required amount of repeat advertising to the same person to induce product purchase. It is widely believed, and probably true, that one advertisement exposure is generally not very effective, and that prolonged campaigns are necessary to change attitudes and behavior. There is no doubt, however, that wide reach--coverage of an audience with at least one advertising exposure--is at least a minimum necessary condition for advertising effectiveness, so reach is universally used as one method for determining advertising effectiveness.

It should be remembered that reach is a measurement derived from vehicle audience size, the size of the audience reached or covered by one or more issues of one or more vehicles.

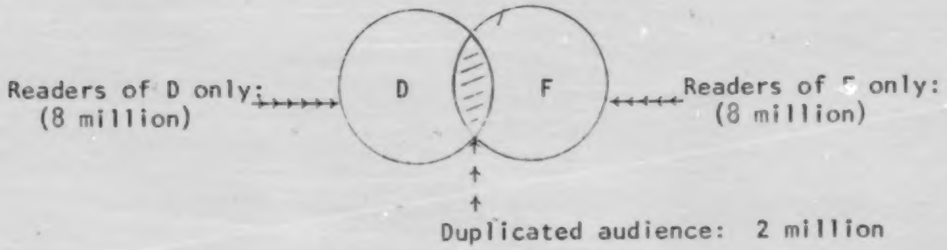
Unduplicated audience. Unduplicated audience, or net audience, refers to the reach of several different vehicles. (The previous concept, accumulative audience, refers to the reach of several issues of the same vehicle.) For any pair of national magazines, some of the same persons will read both and thus be "duplicated"; the persons reading one or both of a pair of magazines are the unduplicated audience. It's obvious that when wide reach is the criterion for media selection, then the objective is to select media combinations or "media mixes" that will provide the largest unduplicated audience. The example below will illustrate this point.

Assume that you wish to select the best pair of magazines out of three available, Magazines D, E and F. All three (for the sake of simplicity in this example) have the same average issue audience, ten million each. Which two will give the widest reach--i.e., unduplicated audience?

Look first at the DE pair. They reach almost entirely the same people, that is they have a high duplicated audience, consequently a low unduplicated audience.



Now look at the DF pair. They reach very few of the same people, that is they have a low duplicated audience, consequently a high unduplicated audience.



The table below sums up the results for the two pairs:

Magazines	Total audience	Duplicated audience	Unduplicated audience
D + E	20 million	9 million	11 million
D + F	20 million	2 million	18 million

The DE magazine mix has an unduplicated audience of 11 million, DF counts 18 million. Obviously, DF is the better media mix in terms of unduplicated audience. Equally clearly, magazine E's audience duplicates the audience for D and adds little to reach.

When a large advertiser wants to place advertising in several magazines to reach as many persons as possible, he thus looks at them in terms of the unduplicated or net audience of various mixes. Just as in the case of the accumulative audience for several issues of a single publication, the audience efficiency or cost-per-thousand-exposures can be computed for combinations of several magazines, by comparing the total cost with the unduplicated audience size.

In a simple example such as the one above involving only three magazines, the computation is simple. In actuality, however, choosing the best combination of five magazines from a list of 100 available can become rather complicated.

Believe it or not, from a list of 100 magazines, there are 75,287,520 different five-magazine combinations. This has led to the increasing use of computers in media research, and the use of such statistical techniques as linear programming, decision programming, incremental analysis, algorithmic analysis and so on. There is not room here to deal with all those complex techniques for determining unduplicated audience and cost efficiency for various media mixes. However, the concepts of those techniques are easy to grasp from the three-magazine example above.

There is a standard syndicated service available to advertisers which simplifies the job of selecting magazine combinations which will maximize unduplicated audience. The Starch Consumer Magazine Reports, published annually and based upon a national survey of magazine reading, provides audience data for 55 magazines and the unduplicated audiences for all pairs of those magazines. From knowledge of unduplicated audience of pairs of magazines, various statistical procedures may be used to estimate the coverage of 3-magazine, 4-magazine or larger combinations.

Measuring Issue Exposures to Print Media Vehicles

Thus far, the discussion has dealt with the audience size, or total number of people exposed to a vehicle or vehicles as a basic measurement of advertising's potential effectiveness. An additional consideration in evaluating media, however, is the number of times a person in the audience exposes himself to a vehicle unit.

For example, take two magazines that are equally efficient in terms of cost-per-thousand-audience; if one magazine is so interesting that each audience member exposes himself twice to each issue, while the other has only one exposure per issue per reader, the former is probably more valuable to the advertiser. Thus, the efficiency of two magazines in delivering exposures can be computed by an extension of the former example.

This can be illustrated in the case of Magazine A and Magazine G, both of which have the same circulation, audience size and audience efficiency.

	Magazine A	Magazine G
Cost, 1-page non-color ad	$35,000	$35,000
Circulation	7,000,000	7,000,000
Readers per copy	3.36	3.36
Issue audience	23,500,000	23,500,000
Exposures per reader	1.0	2.0
Issue exposures	23,500,000	47,000,000
Exposure efficiency (CPM exposures)	$1.49	$ .75

It's obvious that, everything else being equal, Magazine G is probably a better advertising investment than Magazine A, since it provides twice as many exposures for the same cost.

This still leaves unanswered the question, "Just how valuable is the second exposure?" But it has, at the very least, some additional value even though its exact magnitude hasn't been determined.

Just how does the media researcher measure the number of exposures of each audience member to a magazine issue? The previous discussion on audience measurement dealt with the questioning procedure by which readership of an issue is determined. Direct questioning by the "recognition" method will determine the person who has looked into an issue. But can direct questioning reveal how many times he looked into the issue? Does the average person know how many times he has looked into a magazine copy that has been lying around the house for several days or weeks?

The generally accepted answer is that direct questioning in a conventional sample survey probably cannot accurately reveal how many times a person has looked into an issue.

The only accurate measurement possible is through an extremely complicated controlled experimental procedure, in which different copies of the same issue are placed in a household on different days, with several waves of interviews among the same people. The *Saturday Evening Post* was the first to conduct such a study, under the belief that a "reading" magazine such as the *Post* might get more exposures per person than a "looking" magazine such as *Life*. While the whole measurement process was extremely complicated and expensive and conducted among just a few people, the procedure was something like this:

First day:	Copy #1 of issue arrives in respondent's home
Second day:	Interviewer picks up Copy #1 and leaves Copy #2 of same issue
Third day:	Interviewer picks up Copy #2 and leaves Copy #3

etc.

This procedure was followed for several days. Instead of direct questions on reading, however, the experimental copies had been treated with tiny glue spots between certain pages. After the "used" copies were picked up, it was possible to detect if the magazine had been opened at all, and also the exact pages that had been exposed. The broken glue-seals--so tiny that the readers were unaware they were even there--provided the evidence. This study showed that different magazines do indeed differ in the number of times a person exposes himself to them. However, because of the artificial nature of the experimental procedure, it was recognized that people probably behaved differently with the experimental copies than in the normal magazine-reading situation.

A subsequent experiment with a small number of people used the glue-seal method, plus direct questioning by the recognition method, to determine if direct questions would give the same results. It was found that more actual exposure had taken place than was reported by the respondents; that is, for every 100 page-exposures known to have occurred from the broken glue seals, only about 64 were reported under direct questioning. That, however, is an illustration of *advertising* exposure and perception, rather than *vehicle* exposure, and will be dealt with in the next section.

In summary, it may be said that the number of issue exposures is probably a more useful measurement than a simple "nose count" of audience size. But because of the difficulties in obtaining accurate measurement in the real-life situation, and because of the difficulty of interpreting the value of additional exposures by the same person, the concept of issue-exposures is not used widely in evaluating media vehicles. Further research on measurement techniques might someday result in a practical and economical method of measurement, to go along with its undoubted theoretical significance.

Measuring Advertisement Audiences and Exposures in Print Media Vehicles

The previous section dealt with audiences and exposures of whole issues of magazines and newspapers. However, exposure to an issue doesn't insure exposure to the advertising within the issue. As previously noted, the media vehicle must take the responsibility for delivering the audience to specific advertisement space within an issue. A magazine that is highly interesting throughout its pages, that creates a favorable expectation among its audience, that uses layout and format effectively to lead the issue reader to every page, will probably produce more advertisement readers and exposures than will a magazine with an equal issue audience but lesser "editorial appeal." The ability to create advertising exposures is what the advertiser seeks in a vehicle. Once the vehicle has produced that exposure, then the quality of the message determines subsequent impact in terms of perception, communication and behavioral response.

Remember that advertisement exposure doesn't necessarily mean advertisement perception--a person may have been exposed without conscious awareness of the fact, without remembering or changing attitude as a result. Advertisement exposure is simply the opportunity for a message to create perception and communication.

The value of a vehicle that is superior in creating advertising exposure is illustrated in the following example, where efficiency can be computed just as in the previous examples.

Here, we will compare a fictitious "reading" magazine (mostly text) with a fictitious "looking" magazine (mostly pictures). It turns out the reading magazine has a fractional advantage in creating additional advertising exposures, which results in much greater efficiency for the advertiser.

	Magazine D (Mostly text)	Magazine G (Mostly pictures)
Cost, 1-page non-color ad	$35,000	$35,000
Circulation	7,000,000	7,000,000
Issue audience	23,500,000	23,500,000
Average ad exposures per reader	1.4	1.1
Total ad exposures	32,900,000	25,850,000
Ad exposure efficiency (CPM ad exposures)	$1.06	$1.27

The same questions that applied to the previous sections on issue exposure also apply to advertisement exposure. How valuable are the extra ad exposures, and how can one measure ad exposures?

Regarding the value of second exposure, a previously cited Saturday Evening Post experiment concluded that a second exposure is equally or perhaps more valuable than the first in creating brand familiarity and awareness. Logic tells us that it certainly has some additional value, even though the exact magnitude is difficult to measure. Because of the artificial nature of the controlled experiments thus far, the latter conclusion is probably the safest.

A practical method for measuring advertising exposure in the usual sample survey situation has not yet been worked out. Only through the somewhat artificial conditions of the "glue-spot" experimental technique have any valid measurements of exposure been obtained thus far. Nevertheless, that experimental approach has conclusively demonstrated that magazines do differ in their ability to deliver readers to the advertiser's message.

A study sponsored jointly by Reader's Digest and Saturday Evening Post indicates that a monthly magazine issue will be picked up more times during its longer life than will a weekly or bi-weekly magazine and will deliver more advertisement exposures. It also substantiated the belief that "reading" magazines are picked up more times and deliver more advertising exposures than "picture" magazines. These results are shown below:

	Reader's Digest (monthly, text)	Saturday Evening Post (weekly, text)	Life (weekly, pictoral)	Look (biweekly, pictoral)
Issue exposures per reader	5.0	1.9	1.4	1.6
Ad exposures per reader	1.7	1.4	1.0	1.1

About 63% of the ad exposures took place on the first issue exposure, the remainder on subsequent issue exposures. It was further demonstrated that the nature of the advertisement itself did not affect ad exposure, thus verifying that advertisement exposure is indeed a function of the vehicle and not the message: ad exposures was identical for color ads and black-and-white ads; for full-page ads and part-page ads; for appliances, automotive, clothing, food product and insurance ads. (Differences in the advertising message do show up, however, in the measurement of advertising perception, to be discussed later.)

The average number of exposures to an advertisement in the Saturday Evening Post varied according to reader characteristics as follows:

All readers	1.4
Male	1.5
Female	1.3
Age 10-24	1.6
25-39	1.7
40-54	1.1
55 and over	1.2
Some college	1.3
H.S. graduate	1.6
Less than H.S.	1.3
Married	1.5
Not married	1.3

Issue and advertising exposure also varied with certain other demographic characteristics such as income, occupation, geographic region, and so on.

In summary, you may conclude that advertisement exposure is a valuable concept, indicating an important function of the advertising vehicle that has been established expermentally, but no practical method for regular measurement has yet been devised.

Measuring the Print Vehicle's Contribution to Advertisement Perception

In previous discussion it has been emphasized that (a) the media vehicle is responsible for delivering circulation, issue audience and exposures, and advertising audience and exposures, while (b) the quality of the advertising message determines advertising perception and communication. Research on the media vehicle's functions has been labelled media research, while research on the advertisement's functions is called message, or copy, research.

However, there are certain kinds of media research studies which obscure the apparently sharp division between the vehicle's and the message's contributions to overall advertising effectiveness.

The fact of the matter is, the vehicle can contribute to advertising perception not only by providing a larger base of exposures, but by having a reputation for believability and other qualities, and by the context in which an ad appears. The differential value of two magazines in creating advertising perception (conscious awareness of having been exposed to an ad) has been demonstrated experimentally in the following way: an identical ad placed in two different magazines, with conditions controlled so as to produce identical issue exposure and ad exposure, produced different recognition perception scores in the two magazines. (This was a small-scale unpublished study conducted by the Curtis Publishing Company.)

The Advertising Research Foundation has concluded, after a thorough investigation of the various measures of media effectiveness, that a reliable method for measuring the advertising perception value of different media would provide the advertiser with a more useful measure than the prevailing measures of audience or exposure. The method of perception measurement has not yet been worked out to reliably separate the effect of the media vehicle from the advertising message. However, two kinds of studies are currently widely used to give clues to the advertising perception value of different media. They are the reader traffic (or item readership) survey and the media image survey.

Reader traffic surveys. In the same sample survey in which issue audience is determined, questions can be added to find the readership of individual items in the issue. Using the "recognition" method, once a person has been verified as an "issue reader," he can be asked by the interviewer to look at every item in the issue and indicate which of them has been seen or read.

Some magazines and newspapers conduct their own reader traffic surveys, but most rely on the Starch Readership Service and other syndicated commercial research services.

The questioning procedure goes something like this, though it may vary in slight details from one survey to another:

(After the respondent is identified as an "issue reader.")

INTERVIEWER: As I turn the pages of this issue, please let me know which items you have at least seen before.

RESPONDENT: (Pointing to item) I remember seeing that one.

INTERVIEWER: (Pointing to different parts of the ad) Did you happen to see this part (illustration)? This part (headline)? This part (trademark)? This part (printed copy)? How much of the copy did you read?

From these, or similar questions, each advertisement thus receives several scores:

NOTED OR SEEN: The per cent of issue readers who say they had previously seen the advertisement in that particular magazine.

READ PART: The per cent of issue readers who indicate they have read at least part of the written copy in the ad.

READ MOST OR READ ALL: The per cent of issue readers who report reading more than half of the written copy.

In addition to those principal scores, the per cent who say they have read each of the various portions of the ad may be reported--the per cent seeing the illustration, the per cent reading the caption, the per cent reading the trademark; this latter figure is reported by the Starch organization as SEEN/ASSOCIATED, meaning they saw the ad and associated it with the name of the product.

The main perception scores, however, are the NOTED scores and the READ MOST scores.

The NOTED score is generally interpreted as the attention value of the ad's illustration, layout and general eye-catching properties. The READ MOST score is generally interpreted as a measure of the communicative properties of the ad, reading the copy being considered a necessary condition for learning or attitude change or behavioral effect.

This recognition method for measuring the attention paid, or readership of, an advertisement is a measure of advertising perception, not of communication. The knowledge of how many people saw or read an ad tells nothing about whether the ad had any effect on knowledge, attitude, or behavior.

An advertiser might use ad readership scores to test the perception-value of two different vehicles. If the identical ad is placed in two different magazines at the same time, any difference in readership score is clearly due to differences in the vehicles. By determining the ratio between the readership score and audience or exposure data, a more precise estimate of pure perception value is obtained. Take the following fictitious example of two magazines, in which the same advertisement was run at the same time:

	Magazine A	Magazine M
Ad cost	$35,000	$35,000
Circulation	7,000,000	7,000,000
Issue Audience	23,500,000	19,000,000
AD PERCEPTION		
% of issue audience		
...Who saw ad	50%	70%
...Who read most	10%	16%
No. of issue audience		
...Who saw ad	11,750,000	13,300,000
...Who read most	2,350,000	3,200,000
Cost efficiency		
CPM seeing ad	$2.98	$2.63
CPM reading ad	$14.89	$10.94

Thus, while magazine A is more efficient in delivering issue audience, magazine M is more efficient in delivering advertising perceptions, which is after all what the advertiser is more interested in: it matters little how large the issue audience is if the advertisement is not perceived.

One may well ask at this point: How does one know that the respondent accurately reports the items he has seen or read, since with the item-recognition method of obtaining item readership data there is no verification of the respondent's report? Is it not possible that a respondent might exaggerate and claim attention to more items than he has actually seen or read?

There is an experimental study which indicates that item readership scores obtained by the recognition method are accurate indications of what a person has actually read. The recognition method, originated by George Gallup in the early 1930's, had been in use for several years when the Curtis Publishing Company decided it was time to establish its accuracy. They designed and executed what was called the "peek-a-boo" study. A number of persons (one at a time) waiting in an office for an appointment were observed through a "one-way mirror" from directly behind their chair. (A "one-way mirror" is one which appears to be a mirror from the respondent's side, but is transparent from the other side.) A magazine had been "planted" beside the chair in which the subjects were asked to wait. While waiting, most of them leafed through the magazine, and an investigator observed through the one-way mirror the items which the subjects stopped at and the extent to which each was apparently read. A record was kept of the items seen, those seemingly read in part, and those seemingly read entirely. The subjects did not know they were being observed. Several days later, under the pretext of another purpose, the same subjects were questioned in their homes about a number of things, including their readership of items in the test magazine. It was found that their claimed reading of items, by the recognition method, closely matched their actual observed reading behavior; a small number of people reported reading items they hadn't read, and an equal number reported not reading items which they had read; the net result was an accurate representation of item readership, since the two kinds of errors cancelled each other out.

The Starch Readership Service publishes an annual ADNORMS report. These figures are annual averages for all magazine ads surveyed by the Starch organization. These are separated by: magazines in which the ads appeared; size of ad; different product categories; use of color; sex of respondent. Such average item readership scores provide "norms" against which individual ads can be compared, and provide advertisers with rough estimates of the advertising perception value of various vehicles. A cost efficiency ratio is also reported, which makes the vehicle comparison job easier.

Media image studies. The "qualitative value" of a vehicle is derived from those intangible aspects of a vehicle that presumably add to its value to the reader and hence add a "halo effect" to any ads that appear in that vehicle. For example, a magazine with a reputation for honesty and credibility in its editorial content will presumably deliver added credability for the advertising content, over and above sheer exposure.

It seems reasonable that, if two magazines have the same audience and exposure, the one that is more credible will provide a "climate" that will result in greater persuasive effect for an advertisement. Qualitative values are measured by media "image" studies.

Consequently, many surveys are conducted in which respondents are asked to rate or compare magazines on such qualities as:

Attractive appearance
Convenient to read and carry
Writing style
New ideas
Thorough coverage
Modern and up-to-date
Pleasurable to read
Personalized
Stimulating
Fresh and imaginative
Truthful and accurate
Believability of advertising
and so on

The methods for measuring media vehicle "images" are the same as for other kinds of attitude and image measurement described in an earlier chapter.

The results of such studies show clearly that different magazines have clear-cut differences on various qualities; every publication has some qualities on which it is superior to other publications, and some on which it is inferior. The problem in evaluating such data is that no one knows which are important qualities and which are unimportant in contributing to advertising effectiveness.

Take the following example of two magazines which were rated on various qualities in a sample survey, the maximum possible rating being 100.

	Average Rating	
	Magazine P	Magazine Q
Lively	89	72
Dynamic	69	42
Courageous	70	55
Timely	91	62
Informative	90	75
Enjoyable	78	90
Intelligent	90	90
Authoritative	70	70
Reliable	55	65
Ethical	44	69
Friendly	68	89
Thorough	53	75

Magazine P is seen as more lively, dynamic, courageous, timely and informative. Magazine Q is seen as more enjoyable, reliable, ethical, friendly and thorough. They are seen as equally intelligent and authoritative. The question is: which magazine is better in terms of qualitative effects on the reader and the advertising?

Since there is no way of knowing which of these are the more important qualities, little has been added to our knowledge on the effectiveness of each as an advertising vehicle.

The interpretation of media image data is still a puzzle to advertising researchers. This explains to some extent why advertisers must to a large extent still rely on clearly quantifiable data such as audience and issue exposure in evaluating media; while such data do not go as far in measuring media effectiveness as one might wish, they do provide unambiguous numerical data of known usefulness and clear-cut interpretation.

* * * * * * * * * * * * * * *

Many of the concepts used in research on the print media are applicable to the broadcast media, radio and television. However, because of the differences in physical characteristics and in the communications processes, the methods of measuring the concepts of print and broadcast advertising are different. The next chapter deals with the concepts and methods of broadcast media research.

CHAPTER VI. BROADCAST MEDIA RESEARCH METHODS

Radio and television are the major broadcast media. In evaluating their effectiveness, the steps in the advertising process already discussed serve as a model:

- -Broadcast vehicle: a specific broadcast program (e.g., the 6-6:30 PM news program, the Lawrence Welk show, etc.) or time period
- -Vehicle unit: a particular instalment of the broadcast vehicle, or the average instalment
- -Distribution: the number of sets tuned in to the average instalment of the program
- -Program exposures: the number of times each person is exposed to the program, summed up for all persons in the program audience (Not as useful in broadcast as in print media)
- -Advertisement audience: Commercial audience, the number of people exposed to a particular commercial time-period.
- -Advertisement exposure: the number of times a person is exposed to a commercial time period, summed up for all persons in the commercial audience (Not as useful in broadcast as in print media)

Similar methods are used in both TV and radio media research. Because more research has been conducted on TV this section will concentrate on television research, with the understanding that, except where noted, the same methods also apply to radio research.

In some respects, broadcast media research is more complex than printed media research; in other respects, it is simpler.

One complexity arrives in the definition of a television vehicle. A magazine vehicle is clearly identifiable by name, is the same from one place to another, has concrete physical units of distribution which can be counted. Television vehicles, however, may be defined as different programs; or different stations; or different networks; or different time periods or in other ways. In practice, a television vehicle must be defined as a block of time on a television station or network which can be purchased for advertising purposes. Thus, a TV vehicle can range all the way from a single 10-second spot commercial on a single local station to a 39-week series of one-hour programs appearing on all stations of a network.

Another complexity of broadcast media research is in the measurement process. A television message leaves no physical tangible trace that it has entered a household. There is no material "countable" object which can serve as direct evidence of broadcast vehicle distribution, as contrasted with the print media where physical vehicle units are received by the audience. The "recognition" techniques for measuring print audiences are difficult if not impossible for television, since an interviewer cannot usually "show" a previously run program or commercial in obtaining audience and exposure data. In cases where the same commercial appears many times on different broadcast vehicles, there is great possibility of confusion regarding the specific vehicle unit on which the commercial was attended.

In at least one respect, however, measurement of broadcast audiences is simpler than magazine or newspaper audiences--and that's because of the single fixed reception time. A particular instalment of a program appears only once, at a prescribed time, and there is no waiting to allow for it to reach the total audience; the broadcast is received at a particular time or not at all, as contrasted with a monthly magazine where a reader may delay exposure for several days or weeks and accurate measurement requires allowance for a suitable delay in order to get accurate total audience data. Only rarely would a person be exposed more than once to a one-minute commercial time slot.

Because of the impermency and the fixed reception time of broadcast vehicles, much broadcast media measurement is of the *concurrent* type; that is, measurement of distribution and exposure is conducted simultaneously, or almost so, with the broadcast itself. With print media, measurement is *retrospective*, i.e., measurement takes place after exposure.

The principal methods for obtaining broadcast ratings are mechanical recording of sets in use, continuous diaries of viewing behavior kept in the household, and coincidental interviews. Each of these is discussed in some detail in the sections that follow, along with some other less-used types.

Mechanical Recording of Broadcast Reception

"Nielsen rating" is a term known in almost every household to mean the popularity of television and radio programs. Nielsen ratings for programs are based on mechanical recordings of set tuning with a gadget called an Audimeter.

The Audimeter was developed and patented by the A. C. Nielsen Company and was first used to measure radio programs in 1942. The Audimeter is a small box-shaped electronic device connected to a TV or radio set in a household. It records the exact times when a set is turned on and off and the station(s) or channel(s) to which the set was tuned during those times. The information is permanently recorded in a removable cartridge installed in the Audimeter; at specified periods, the household head sends the used cartridge to Nielsen headquarters and inserts a new one.

From the continuous record of set operation provided by the Audimeter, one can determine whether and when the set was tuned before a particular program, whether the set was switched before a program ended and whether the set was tuned in during commercial periods. It can detect the flow into and out of any given time period or program.

As with many other forms of advertising research, Nielsen ratings are based on a sample of U.S. households, the sample chosen to be as nearly representative of the total TV-owning population as possible. The sample size of about 1200 Audimeter households may seem small to the non-researchers, when projected as representative of approximately sixty million U.S. households. However, the laws of statistical probability provide 95% assurance that such program ratings are accurate within about 3%. That is, a program with a rating of 50% based on a sample of 1200 households almost certainly (95% confidence) means that between 47% and 53% of U.S. homes were tuned in. This "error" of plus or minus 3% could be reduced to about 1% by using a sample of about seven thousand homes, but the increased precision is not generally considered worth the six-fold increase in costs.

One important point to keep in mind regarding the "Nielsen rating" is that it is a measure of vehicle distribution or circulation and not a measure of audience or exposure. It is a highly accurate measure of the number of sets turned on and tuned in to a particular program, but it does not indicate how many persons, if any, are in the audience or actually exposed. A set in use may have several persons paying attention, or none. As such, the Nielsen rating corresponds to the number of copies of a magazine issue circulated, not to the audience of a magazine issue. In short, it is a precise measure of the first step in the advertising process--distribution--and only through the use of additional methods can the size of the broadcast audience be determined.

The Nielsen figures are generally referred to as *audience* figures but it must be kept in mind that this is merely a convenience term, not corresponding to the precise definition of audience (number of persons exposed) used elsewhere. That limitation must be kept in mind in discussing three separate program measures obtained from the national Nielsen TV rating figures--total audience, average audience and share of audience.

The "Nielsen *total audience*" rating of a program is the number, and percentage, of all TV homes tuned in for at least six minutes.

The "Nielsen *average audience*" rating is the average number of persons tuned in per minute throughout the program, and is equivalent to the average number of homes tuned in through the whole broadcast.

The "Nielsen *share of audience*" rating is perhaps most important to the advertiser in determining the popularity of his program compared with all competing programs on the air at the same time. It is computed by dividing the program's total audience by the total number of sets turned on at the same time.

The following fictitious example will clarify these three measures. The national ratings for all three networks during the 9:30-10:00 p.m. period on Thursday, June 17 are as follows:

Projections from Nielsen Sample
Total TV Households: 60,000,000
9:30-10 pm, June 17, sets-in-use: 40,000,000

Network	Total "Audience"		Average "Audience"		Share of "Audience"	
	Number	Rating	Number	Rating	Rating	
ABC	22,000,000	37%	15,000,000	25%	(37/90)	41%
CBS	18,000,000	30%	13,000,000	22%	(30/90)	33%
NBC	14,000,000	23%	12,000,000	20%	(23/90)	22%
Combined	54,000,000	90%	40,000,000	67%	--	--
Base	60,000,000	100%	60,000,000	100%	54,000,000	100%

Note that the combined total "audience" is 54 million, the combined average "audience" is 40 million. This is caused by some people switching from one network to another (for at least six minutes with each) during the 9:30-10 period and being counted twice in total audience. The combined average audience must equal the total sets-in-use during the period, 40 million.

From these fictitious figures, it is apparent that average audience and total audience ratings give somewhat different evaluations of the programs. While NBC had the smallest total audience by far, it was successful in keeping viewers tuned in for the whole program, or for long periods of the program, and therefore it has a proportionately higher average audience. It's conceivable that one network might have the highest total audience and the lowest average audience, and vice versa.

Average audience is almost always lower than total audience, though a program which "hooked" all of its total audience to watch the entire program would have the same rating for total and for average audience.

Despite its accuracy as a measure of distribution, the mechanical reception recorder has some disadvantages and limitations. It provides no audience or exposure data; the mechanical device is subject to occasional malfunction; people forget to take out the old cartridge and insert a new one; they sometimes neglect to mail in the used cartridge; installation and servicing of the Audimeter may condition or "contaminate" the normal viewing and listening behavior of the household; permission must be secured to install the devices in the first place, and those households that agree may not be truly representative of those that refuse to take part. The latter point--securing cooperation of a truly representative sample of households--is probably the biggest single problem connected with mechanical reception recording.

Instantaneous mechanical measurement. One other problem associated with mechanical devices is the delay in reporting the results to the central tabulation office--cartridges are generally mailed in and replaced every two weeks, with a corresponding lag in reporting results to the advertiser.

The desire for immediate information on program reception has led to several variations on the mechanical measurement method.

A limited number of sets in metropolitan areas can be connected by direct wire to a central point; set tuning is detected the instant it occurs and automatic tabulation makes it possible to obtain ratings from a single city continuously, while a program is on the air. Reports from several cities can be combined, tabulated and reported within a few hours after a program has left the air. This centralized instantaneous measurement system has the advantages of speed and quick detection of mechanical malfunction; however, reports can be obtained only from metropolitan areas with no sample representation from small towns or rural areas, and the cost is relatively high. Both Nielsen, with its "Instantaneous Audimeter," and the American Research Bureau, with its "Arbitron," supply this service.

Mobile detection units are the newest wrinkles in obtaining instantaneous broadcast ratings. A truck loaded with electronic detection gear can pick up electronic impulses while cruising the streets and determine which sets in the vicinity are tuned to which stations, with the results being immediately and permanently recorded on tape. One such method, is an electronic Peeping Tom known as "TEST" (Tanner Electronic Survey Tabulation) demonstrated in Chicago.

A news item also reported that airplanes can be similarly equipped with electronic detection gear and gather TV rating data over a wide area quickly.

Not too much is known at present about the validity of these mobile detection methods, but they hold promise of overcoming some limitations of other rating methods.

Diary Records of Broadcast Reception

A carefully kept diary of broadcast reception can provide data on both vehicle distribution and vehicle audience exposure. Both Nielsen and the American Research Bureau have a continuous panel of diary-keeping households to supplement the mechanical measurement system.

A well-designed diary record form will contain data on the name of the program, the station or channel, the time and date of broadcast, and the time spent watching the program by every member of the household. For accuracy, the diary should be filled in simultaneously with reception, not delayed until the end of the day or later in the week. A mechanical device used by Nielsen to assure prompt diary entries consists of a light-and-buzzer system attached to the TV set which reminds the viewer every half-hour to fill in his diary.

Even so, the diary method has several limitations which make it a supplementary device to the mechanical method and not an adequate replacement. People forget to record what they have heard or seen, particularly in the case of radio listening which is accompanied by other competing activity; shorter and less popular programs are likely to be ignored by the diary keeper while viewing of the most popular programs tends to be exaggerated in the record-keeping process; the households which agree to keep diaries may not be truly representative of those which refuse to cooperate; too much awareness of the record-keeping process probably distorts normal behavior to some extent; accurate recording becomes difficult when there is more than one radio or TV set in the household.

In spite of its limitations, the diary method can provide data from people and places where the mechanical devices cannot be used, and it can provide valuable data on audience exposure and audience characteristics which is unobtainable otherwise. For this reason, the diary method is widely used as a supplement to the mechanical system.

All-media diary records. One of the principal problems involved in comparing broadcast media with print media is that different research methods are customarily used, and different criteria for circulation and exposure are necessary because of the nature of the media themselves. One way out of this dilemma may lie in having a person keep a single diary in which he would record exposure to all media--print and broadcast--in a comparable manner. The Advertising Research Foundation has evaluated a pilot study of this measurement method, the ARB-RKO study conducted in Detroit in 1965. Besides the comparison of different media obtained in the diary, this study also compares diary data with that obtained by telephone coincidental, telephone recall and personal interview methods, thus providing a valuable methodological comparison.

Coincidental Telephone Interviews on Broadcast Reception

An army of interviewers is scattered randomly around the U.S., each of them continuously telephoning a randomly selected sample of telephone homes in his respective vicinity while TV and radio programs are in progress; that's a brief description of the coincidental interview technique. It provides an *average audience* rating based on the program being seen or heard at the instant the call is made; by combining results from all over the country, an average minute-by-minute audience rating is obtained on every program on the air during the interviewing period.

From this basic information, the *share of average audience* can also be figured, along with data on the number and kinds of persons exposed in each household.

Thus, the coincidental method provides measures of both distribution and audience exposure, but cannot provide data on the *total* audience for a program nor on audience "flow" from program to program.

Coincidental ratings were first provided by the Hooper survey company in 1934--as the well-known radio "Hooperatings"--and since the advent of television a similar TV service has been furnished by Trendex.

Some of the advantages of coincidental measurement of broadcast reception are speedy tabulation and reporting, inexpensive cost, avoidance of reliance on memory, almost universal telephone ownership, low interview refusal rate, and no "conditioning" of the respondents by repeated data collection from the same people as in the case of mechanical or diary methods.

There are disadvantages of the telephone coincidental method: its inability to measure total audience and audience flow; some undependability in personal oral reports of program reception (including some exaggeration of attendance); the problem of multiple receivers in the household; the necessity to keep the telephone interviews brief (usually about one minute long); a growing proportion of unlisted telephone numbers.

The advantages of this method far outweigh its deficiences, however, for certain kinds of program reception data. One particular usage for which it is especially well suited is in quick ratings of competitive network program popularity. Network TV executives are particularly interested in the popularity of their own programs compared with those of competing networks. In cities where all networks can be received, the relative popularity of a network program competing against other network programs can be quickly ascertained and tabulated.

Other Ways for Measuring Broadcast Reception

The preceding discussion on the three principal methods of broadcast media research--mechanical recording, diary, and coincidental telephone--accounts for the major part of such measurement. But mention should also be made of other measurement methods used in broadcast media studies--personal coincidental, personal roster recall, personal unaided recall, telephone unaided recall, and various combinations of those methods.

Personal coincidental. This is like the telephone coincidental method, except that face-to-face personal interviews are made. Throughout a given program or time period, respondents are questioned about program exposure at the moment the interviewer arrived.

Personal roster recall. In this aided recall method utilizing personal interviews, respondents are shown a list of programs and stations and asked to indicate which they were exposed to during a previous prescribed time span.

Personal unaided recall. This differs from the roster aided recall method only in that no list of programs or stations is shown, depending entirely upon the unaided memory for exposure information.

Telephone unaided recall. This is different from the personal unaided recall only in that the interviews are conducted by telephone.

Combination telephone coincidental-and-recall. Respondents are asked in a telephone interview about both coincidental and previous broadcast exposure.

Combination telephone coincidental and diary. This method combines broadcast exposure information obtained by the coincidental telephone method in one sample of homes with information obtained by the diary method in another sample of homes.

Combination telephone coincidental and personal roster recall. This method combines broadcast exposure information obtained by the telephone coincidental method in one sample of homes with information obtained by personal roster recall in another sample of homes.

Combination mechanical recorder and diary. This combination is self-explanatory, using two different samples.

Recommended Standards for Broadcast Media Research

If the reader is confused at this point by the great variety of methods for measuring program distribution and exposure, welcome to the club. Some methods measure one thing, some measure another; when they do measure the same thing, they tend to give different results because of the differing techniques employed. Confused by the welter of conflicting results and techniques, broadcasting and advertising executives cooperated with the Advertising Research Foundation in the 1950's to study the problem and come up with some recommended standards against which the various methods could be evaluated.

After intensive study, the ARF's Radio-TV Ratings Review Committee came up with a set of 22 ideal guidelines for broadcast media measurement. None of the methods previously described fills the whole prescription, but some of them come closer than others. The seven principal recommendations are as follows:

1. Broadcast reception should be measured in terms of set tuning (vehicle distribution or circulation), not in terms of persons exposed to the turned-on set (vehicle audience or exposure). While it would be extremely desirable to know about actual attention to the set or audience exposure, no satisfactory way of defining or measuring exposure has yet been formulated. Of all the various levels of exposure that could be used, set tuning or program circulation is the most objective, simplest and most understandable.

2. The unit of measurement should be the household, not the number of sets or number of individuals. "Number of households" has wider application for several reasons, though it is recognized that the number of sets does not equal the number of individuals.

3. All sets owned by the household should be measured.

4. The entire reception area of the station(s) carrying the program should be measured.

5. The measurement should be representative of all households within the reception area.

6. The measurement should report the average instantaneous audience rather than total program audience. The average audience figure automatically weights households in proportion to the amount of time spent with a program, permitting uniform comparisons of programs of different duration. A total audience measurement, on the other hand, counts all households equally regardless of how long they are exposed and for this reason is unsuitable for comparing programs of different duration.

7. The measurement should express the number of households reached. Program measurement can be expressed either as numbers or as percentages of some total; because different bases may be used in calculating percentages, confusion in comparing different programs may result. Reporting in terms of absolute numbers of households tuned in eliminates this source of variation.

In addition to the seven principal recommendations above, representing the basic information that should be reported in every broadcast media study, certain other data were recommended as supplementary information. The following 15 items, not to be regularly reported, pertain to useful but non-essential data on broadcast reception:

8. Total household audience;

9. Unduplicated household audience to two or more broadcasts;

10. Program audience size measurements as percent of all households;

11. Households using any receiver by 15-minute time periods;

12. Audience characteristics, such as sex, age, product usage, income, etc.;

13. Program audience size measurements for specific segments of a program, such as commercials;

14. Full network audience by specific time periods;

15. Audience figures should be based upon a probability (random) sample;

16. Audience figures should be based upon the average daily program;

17. The measurement should be reported for each commercially sponsored segment of the broadcast--that is, each sponsor of a multi-sponsor program should have audience information for his segment;

18. All broadcast hours from 6:00 a.m. to midnight should be measured;

19. The measurement should be available at varying frequencies, depending upon the importance of the market;

20. The measurement should be available within one month of the last measured broadcast;

21. There should be adequate control of sampling errors, by rigorous specification of the design and size of samples--a random sample of 400 is minimum for local measurements, 1,200 for national measurements, during each 15-minute program period.

22. The net effect of non-sampling errors should not exceed the sampling error--non-sampling errors including such things as non-completion of sample, conditioning of tuning behavior, interviewer or reporting errors, and data-processing errors.

It has been mentioned before, but is worthy of reiteration, that every reference to broadcast audience above really means circulation or distribution of sets-tuned-in. This is one of the sources of confusion in media measurement--the term "audience" as used in broadcast media research doesn't mean the same thing as when used in print media research. The chapter on "The Advertising Process" is a useful reference to keep the distinctions between distribution and audience clear.

Evaluation of Different Measurement Methods Against Recommended Standards

With the seven principal standards of broadcast media measurement as guidelines, one can compare and evaluate the 11 principal methods.

Only three of the methods--combination-telephone-coincidental-and-diary, combination-telephone-coincidental-and-personal-roster-recall and combination-mechanical-and-diary--satisfy all seven standards for both radio and television media measurement.

The deficiencies of the other methods are as follows:

Measurement method	Deficiency: Doesn't measure
Diary	Average instantaneous audience
Personal roster recall*	Same as above
Personal unaided recall	Same as above
Telephone recall*	Same as above
Mechanical recorder	All sets in household--can't measure out-of-home sets in use (portable, car, etc.)
Personal coincidental	Same as above
Combination tel-coincidental & tel-recall*	Same as above
Tel-coincidental*	Same as above

*(Telephone methods have an additional limitation: they are representative of all households only to the extent that telephone ownership approaches saturation.)

It will be noted that no single method satisfies all criteria--only the combined methods provide all the data needed in adequate broadcast measurement. This illustrates a point that occurs time and again in any form of advertising research--a variety of research methods and techniques is usually needed to give a true picture of whatever situation is being investigated.

One other point should be made in closing out this chapter on broadcast media research. The mobile-mechanical method is still in the developmental stage at this writing; if it can be perfected, it could eliminate many of the problems associated with the currently available methods of measuring broadcast distribution or circulation. This still leaves unresolved, however, the problem of measuring individual audience members exposed to a broadcast--a problem for which there is no immediate prospect of solution. For the same reason, advertising exposure is an unknown quantity in broadcast media research.

CHAPTER VII. OUTDOOR MEDIA RESEARCH METHODS

The principal forms of outdoor media are posters and car-cards. Posters are the familiar billboards seen mounted along streets and highways, sometimes called "panels." Car-cards refer to the ads carried on the outside of buses and other public vehicles, sometimes called "outdoor transit posters" or "transportation advertisements."

This is a relatively neglected medium from the standpoint of research This is possibly due to the fact that outdoor advertising represents only one or two per cent of total advertising expenditures in this country.

Because of the dearth of published outdoor media research studies, this discussion will necessarily be brief and confined principally to posters.

Due to the peculiar physical nature of the medium, outdoor research methods are somewhat different from those used in other media. However, the familiar advertising process concepts of vehicle distribution, vehicle exposure and advertising exposure will be useful in understanding the basic principles of poster measurement.

In the poster medium, a vehicle may be considered as all poster locations (billboards, etc.) available for advertising purposes from a single poster firm within a single city or geographic area. (This may be compared with the magazine medium, in which a vehicle is a single magazine title, such as Reader's Digest or Saturday Evening Post.)

A poster vehicle or "buy" is generally available to advertisers in 30-day segments, thus defining the vehicle unit as a 30-day showing of posters within a single geographic area. (In the magazine medium, the vehicle unit is defined as one issue.)

Vehicle distribution for posters is defined as the number of posters available or used in a given geographic area. In practice, this is translated into the per cent of coverage of major traffic routes in an area. For example, a "100 showing" is the number of properly located posters required to establish complete 100% coverage of major traffic routes in an area; a "100 showing" in a large area such as Los Angeles might require 400 posters, whereas in a smaller area such as St. Louis only 150 posters might be required for a "100 showing." (In the magazine medium, vehicle distribution refers to the number of copies circulated for a specific issue of a specific magazine.)

Vehicle exposure for posters, as in other media, is defined as the number of people who are known to have the opportunity to see the poster. This means that anyone travelling along an appropriate route in the proper direction who passes within range of a poster showing is counted as a member of the vehicle audience, equivalent to one exposure. This is usually translated into a percentage of the total area population, yielding a "per cent coverage" figure. (In the magazine medium, vehicle exposure is equivalent to the number of people who have seen or looked into a copy of a particular issue.)

The Traffic Audit Bureau is a national research organization which provides reports on the number of people and number of vehicles passing within range of poster locations. This is useful advace information for the advertiser, telling him the amount of exposure he an expect to obtain with any size showing in a particular area. The Traffi Audit Bureau's function for poster advertising is quite similar to that served by the Audit Bureau of Circulation (ABC) for the magazine and newspaper media; one difference is that ABC provides data on vehicle distribution (circulation) while the TAB is concerned with vehicle exposure.

One interesting difference in the poster advertising process, as compared with other media, is that *poster vehicle exposure is the same as poster advertising exposure*. The poster is both the message and the carrier of the message; it's impossible to be exposed to one without being exposed to the other. This may be contrasted with TV and radio commercials which are only part of a larger vehicular context, or with magazine and newspaper advertisements which are only part of a copy. In both broadcast and print media, the audience pays deliberate attention to the vehicle out of personal choice; in the case of posters, exposure to the vehicle and the advertisement is purely incidental, or accidental, to other purposes of the audience, and this probably affects the way in which the message is perceived and communicates.

What methods are used for actually conducting poster research on vehicle distribution and vehicle/advertising exposure?

Poster distribution is, of course, relatively easy to determine simply by driving along prescribed routes in an area and counting the number of poster locations. This is far simpler than for other kinds of media and in practice the poster company records are generally accepted as correct, though some verification may occasionally be made by the advertiser or his agency.

Poster exposure data is obtainable by four methods--the traffic-count, the travel survey, the recognition survey, and the photographic survey.

The traffic-count survey of poster exposure is conducted by counting the number of persons walking or riding in the proper direction past the poster locations. Previous studies have shown that the mode of travel has a direct bearing on the probability of exposure; therefore, pedestrians and people in passenger cars are discounted 50 per cent to allow for non-viewing, while bus passengers are discounted 75 per cent. This sort of survey research gives reasonably accurate average estimates of exposure by taking random samplings of time-of-day, at random samplings of location along the different routes within an area, on different days of the week and at different seasons of the year.

The table below, based on a Nielsen survey conducted in Los Angeles, shows the relationship between size of showing, number of posters, and poster exposure for different durations of showing.

Los Angeles Poster Exposure (% population coverage)

Size of showing	Number of posters	Cumulative unduplicated audience	
		After one week	After four weeks
10	38	34%	63%
15	57	49	77
25	107	64	90
50	200	75	95
75	305	84	97
100	400	90	98

Note that the size of showing is directly related to the number of posters, except for minor variations in the smaller showings. It's also apparent that a given desired coverage can be achieved with a small showing for a longer period of time or with a large showing for a short period; for example, 90% coverage can be obtained with a No. 25 showing over a four-week period, or with a No. 100 showing in one week.

One advantage of poster advertising is its ability to achieve almost complete coverage of a given area. A No. 100 showing for four weeks duration will be exposed to 98% of the adult population in the Los Angeles area. Probably no other single advertising vehicle can approach this figure--not even the newspaper or magazine with the highest circulation in an area, nor the highest rated television or radio program. On the other hand, the perception and communication rate of posters is probably considerably lower for a given number of exposures than in the other media. An exposure, remember, simply means someone within range of a given vehicle. A certain amount of deliberate attention and motivation to perceive is present with newspapers, magazines, radio and television. Attention to posters, on the other hand, is almost purely incidental and non-deliberate. Another detracting feature about the communication power of a poster is the extremely brief nature of exposure by passing traffic; the message must be briefer and more immediately comprehensible than in the other media. From the standpoint of research, this means that recall, information gain, and attitude change from a single exposure to a poster message are very slight, resulting in a knotty problem for reliable research measurement.

The travel survey method of measuring poster exposure is accomplished by selecting a random sample of individuals in an area, then questioning them intensively on all their local travelling, by foot or car, during the preceding week. Since the location of all posters in the area is also known, the journeys can be related to the poster locations and poster audiences estimated therefrom. The travel survey relies on accurate sampling and is highly subject to memory errors and is therefore a more indirect and less accurate measure of exposure than the traffic survey. It does, however, permit gathering information about respondent characteristics, and other interview-obtainable data.

The recognition survey method operates by showing, to a sample of respondents in an area, reproductions of outdoor posters that have appeared and asking which have been seen. An adjustment for mistaken exposure claims can be made by including some poster designs that have not appeared in the area. In practice, recognition data can be obtained simply by adding the appropriate poster reproductions and questions to a travel survey interview. It's difficult to determine whether a recognition survey is really measuring vehicle exposure, advertising exposure or advertising perception; the same data with variations, are used for all three measures.

The photographic survey method has not been used extensively, but it is at least theoretically the best method of measuring poster exposure. One demonstration of this method was conducted by Alfred Politz Research. Cameras facing the highway were concealed in a poster location and photographs of passersby were taken automatically at selected time intervals. Special lighting and filters were required to snap the night-time traffic. The photographic method has the advantage of not relying on respondent memory (as in the case of the travel survey) nor on the tallying accuracy of human recorders (as in the traffic survey). However, the installation, maintenance and expense of camera equipment make this method prohibitive for all except special studies.

* * * * * * * * * * * *

There has been even less research on car-card or transportation advertising than on posters. However, in passing it should be mentioned that all four poster audience measurement methods--traffic survey, travel survey, recognition survey, and photographic survey--are equally well suited for measuring car-card audiences.

CHAPTER VIII. MESSAGE RESEARCH IN GENERAL

The message itself is the final link in a chain of elements, all of which must be strong in order for advertising to be effective.

Perhaps it will be helpful to repeat here some of the previous links and the choices based on research that should have already been made prior to the message decisions. These links in the advertising chain include:

1. Market and product. The product should be something for which a need exists in the marketplace; or, in the case of a radically new kind of product, the public must be "educated" regarding their need for it. The product must also be of sufficiently good quality relative to the other choices available to the public. The best advertising cannot bring lasting success to a poor product, or one for which no need exists. "Product research" also includes work on the product name, price, package and consumer motivations; all this is not, strictly speaking, advertising research though some of the information developed in product research may be useful in planning advertising.

2. Money for advertising. Given a good and needed product, too little money for advertising can slow down sales, too much will result in financial waste.

3. Media. As just discussed in detail, this means choosing media and vehicles for maximum vehicle distribution, vehicle exposure, advertising exposure and to some extent, advertising perception.

4. The final step, Messages. This requires both choice and execution of advertising messages. In terms of the advertising process model, message research is concerned with enhancing advertising perception and communication and certain behaviors, including purchase.

Message research includes both pre-testing (also called copy research or creative research) and post-testing, for all varieties of media, in all stages of preparation from the idea or concept to the finished execution.

How NOT to Determine Ad Effectiveness

Before going any further, a large dose of cold water should be thrown on one primitive but still popular method for judging advertisement effectiveness.

This might be called the "seat-of-the-pants" method. Sometimes it takes the form of having an "expert" or critic to look through a batch of ads or commercials and pass judgment. Sometimes it takes on some psuedo-scientific trappings, as by getting the opinions of a group of experts who are asked to judge the quality of the advertising.

Perhaps the best criticism of this method is provided by simply quoting from an expensive book entitled The Best Advertisements from Readers' Digest 1955-1961. Here are the author's comments on how he selected several dozen ads from 3,387 that had appeared in a seven-year period.

> So that leads us to the question that started me writing this preface in the first place: How did I pick the ads in this book?
>
> When the assignment began, I sent a three-page letter to several hundred advertising leaders outlining the project and asking for nominations.
>
> Several hundred exceedingly helpful replies resulted. You would be interested, I think, in reading the lot.
>
> Most of my choices were also nominated by others, so the book has a solid core of almost unanimous professional opinion, unprejudiced and unbiased.

> I grew prouder and prouder of this advertising business as I read letter after letter in which my "judges" paid no attention to whether or not their own companies had anything to do with the ads they nominated. They just called them as they saw them.
>
> Other choices are mine alone, without benefit of readership figures, result stories or anything else beyond my more than forty years' experience as a working creative advertising man.

While subjective judgments of this sort can be a useful guide to the technical production quality of ads, several basic research studies have been conducted that point to one overriding fact: Experts in any field don't see or evaluate things in the same way as the public. Art critics, drama critics, or advertising critics--all are looking and evaluating on different grounds than the public for whom the product is intended. Another crucial point is that any group of advertising experts will disagree among themselves when asked to choose the most effective advertisements from a larger batch. And there is no denying that a great many advertisements--presumably having already passed the "test" of expert opinion--are resounding failures when circulated in the real-world.

This is not to say that creative skill and judgment of writers, artists, and producers can be dispensed with--far from it. But it does indicate that the crucial test of advertising is whether it has the intended effect on consumers, usually the improvement of attitudes or sales.

Quite often an informed opinion is necessary; not every decision has evidence available to back it up, and there is often not time to do research that would yield the best answer to a problem. However, there are many cases where research evidence is available, or can be obtained by sufficient forward planning. When possible to use research, it is inexcusable to base highly expensive advertising decisions on anything other than the effects of the advertising on consumer behavior.

The Extent of Message Research in Advertising

Farther on will be found a listing of typical studies in various aspects of message research. New methods and additional studies are constantly in progress in advertising agencies, commercial firms and universities. Most of these studies are never published, since they deal with confidential matters paid for and used by clients.

Most people may assume that all the tremendous flood of advertisements in newspapers, magazines, radio and television are researched before being distributed. Such an assumption is largely wrong, even for the biggest advertisers. While almost all advertising agencies do some message research, by no means is every advertisement subjected to such scrutiny. Consequently, much of the advertising that reaches the public is irritating, in poor taste or just plain ineffective. One commercial research firm has estimated that only one out of thirteen commercials tested achieves its purpose. And there is even some direct evidence, from post-testing, that advertising can have a negative effect on sales; that is, advertising caused people to think less of the product and actually prevented some sales that would have occurred without the advertising.

The only insurance available to an advertiser, to insure against ineffective or negative-effect advertising, is pre-testing. When it is considered that a single magazine or television ad can cost $100,000 or more, it's only common sense to avoid expensive mistakes by investing a small fraction of that amount in message research. One large advertiser is still trying to forget about a recent year-long advertising campaign costing $60,000,000 in all media which had absolutely no effect on product sales except to drive many customers to the leading competitor.

Some Questions Involved in Message Research

Regardless of whether an advertisement is intended for the print media, the broadcast media or outdoor media, there are several basic general questions by the decision-maker which can only be answered by message research.

These may be generally phrased as follows:

1. Which selling concept should the advertising emphasize?
2. Which theme best expresses the concept?
3. What's the best graphic approach to maximize perception?
4. Which persuasion strategy will maximize communication?
5. How should the body copy be written?
6. How effective is the complete execution of the finished ad?

Those basic questions can be elaborated upon as follows.

Out of all the things that might be said about a product, which concepts should the advertising concentrate on? In the case of an automobile, for example, should the advertising talk about beauty, durability, comfort, speed, economy or what? In other words, what broad concepts or "sales propositions" or "sales ideas" should be stressed out of all those pertinent to any given product and brand?

Given an effective selling concept or sales proposition, what's the best specific theme to use in a particular advertisement? For example, one can discuss the "economy" selling concept in terms of initial price, low maintenance cost, long-life and in several other ways. Should the same theme be used throughout a campaign, or should many different themes be used to express a single selling concept? Which ones?

What is the best graphic approach for attracting attention and enhancing perception? In a magazine vehicle, should an illustration be used or not? If so, what kind of illustration? Artwork or photograph? Black-and-white or color? How much white space? What kind of typeface? How much body copy? Or, in the case of a television commercial, what proportions of visual and spoken are most effective? Should it use live action or cartoon treatment? Should there be a music background? If so, should it be soothing or irritating? One continuous scene or several different scenes?

What's the best persuasion strategy to use? Is it better to use "two-sided" messages in which the competition is mentioned, or a one-sided approach? Should it be an emotional approach or a rational approach? If an emotional approach, should one use fear, humor, pathos or what? Should one talk about product features or the consumer benefits derived from those features? Hard sell or soft sell? Presented by a famous person, an anonymous announcer, or an "off-camera" voice? If a famous person, which one?

In determing how the body copy or script should be written, it's necessary to consider such things as sentence length, word difficulty, special words for special kinds of customers, creation of abstract or concrete images, loose or tight writing, regional colloquialisms and slang, and various other aspects of the creative art which are difficult to specify.

Finally comes the crucial question: how does the total execution of the finished or semi-finished ad hold up? If the individual elements of the message--the selling concept, the theme, the graphic approach, the persuasion strategy, the body copy--are all demonstrably effective by themselves, chances are that the complete ad execution will be effective. But there's always the possibility of achieving an "oil-and-water" combination of elements which, for some reason, don't mix. Only a test of the finished execution, or of alternative versions of the execution, will provide maximum insurance against a flop. This is a form of "quality control" which should be applied whether or not all specific elements of a particular ad or campaign have been previously tested.

It may not be necessary to conduct research on all these questions for every specific advertisement or commercial, but somewhere along the line all those questions must be at least considered. In the opinion of many professionals, research on the concepts and themes of advertising is of greater urgency than research at later stages of message development.

Typical Message Research Studies

Some of the published articles on advertising and communications research provide a good overview of interest-areas in message research. The articles mentioned below, mostly from the Journal of Advertising Research and from Journalism Quarterly, will illustrate some attempts to answer through research the general questions the advertising decision-maker needs to know.

In the realm of research on selling concepts, Fisk's article on "Price Rivalry Among Philadelphia Food Chains" determined through content analysis the different concepts (message strategies, selling propositions) used in competing grocery advertisements. Mukherjee's article "A Factor Analysis of Some Qualitative Attributes of Coffee" dealt with how he had consumers to drink different brands of coffee, and rate them on 14 different attributes (taste, smell, etc.); then, through the statistical technique of factor analysis, he determined which coffee attributes were most important to use as advertising concepts. Axelrod's article "Reducing Advertising Failures by Concept Testing" stresses the importance of this phase of message research.

Research on advertising themes is the topic of Henderson's article "Sales Effects of Two Campaign Themes"; this was a controlled field experiment of complete ads under realistic conditions, with sales as the dependent variable. Vickers' article "A Pre-Test of Four Australian Paint Ads" also used complete executions to select themes on the basis of attitude and opinion responses of consumers. Haskins' article "Pre-Testing Editorial Items and Ideas for Reader Interest" describes a method by which article and advertisement themes written as headlines can be rated for reading interest as a basis for theme selection. Felstehausen's article "Headline Motivation and Reader Response in a Farm Magazine" takes a similar approach to theme selection. Swanson's article "What They Read in 130 Daily Newspapers" determined the reader interest in various subjects or themes by computer analysis of readership studies; while his article deals with news stories, the same technique has been used in analyzing advertisement readership.

Probably the largest body of published message research deals with graphics or "presentation techniques." Belk's article "Immediate Recall of TV Commercial Elements" analyzes responses to a number of commercials to determine the relative effects of picture, print, voice, sound and length. Greenberg's article "Visual Materials and Recall of Magazine Articles" compared the effects of magazine items with and without accompanying pictures. Robinson's article "How an Advertisement's Size Affects Responses to It" tells how the eye-movement camera was useful in evaluating responses to different-sized ads. Troldahl's article "Predictors of Newspaper Advertisement Readership" compared the relative effects of creative treatment, ad size, and product by analyzing readership scores of a large number of ads. Wells' article "Communicating With Children" showed how TV commercials involving motion were more effective with children than if "static" presentation were used. Ramond's article "Scaling Visual Dimensions of Advertising" describes how laboratory visual equipment was used to find relationships between various ad elements and the viewer's responses. Typography was investigated in Haskins' article "Testing Suitability of Typefaces for Editorial Subject-Matter," in English's article on "Use of Split-Run Techniques in Studying and Typography," and in Wrolstad's "Adult Preferences in Typography." The use of pictorial matter was the subject of Kerrick's article "The Influence of Captions on Picture Interpretation" and MacLeans' article "Picture Selection: An Editorial Game." Tinker's article "Effect of Line Width and Leading on Readability of Newspaper Type" is representative of another much-investigated area of graphic research.

The use of different persuasion strategies has been the topic of a great deal of basic psychological research published mostly in various psychological journals; most of these reports are on non-advertising topics such as politics and social problems. Advertisers have been slow to utilize the findings of that research, or to do research of their own in this very important area. Kover's article "Selecting Commercial

Spokesmen" is an exception, illustrating research on the necessary characteristics of a persuasive spokesman. Book's article "Recall of Institutional TV Commercials" reports on differences in the persuasion strategies to be used for women and for men viewers. Adams' article deals with "The Relative Credibility of 20 Unnamed News Sources," illustrating that message believability can be improved by use of the proper message spokesman. Crane's article "Immunization--With and Without the Use of Counter-Arguments" investigates ways to "innoculate" people against later competitive messages.

Research on copy writing techniques deals with both headlines and body copy. Ways to improve understandability or comprehension was the subject of Goldsmith's article "Comprehensibility of Initials in Headlines" and Kearl's "Estimating Understanding of Scientific Terms." Another requirement of good copy is readability, as dealt with in Carter's article "Cross-Cultural Application of Four Flesch Formulas," Taylor's "Cloze Procedure: A New Tool for Measuring Readability," and Haskins' article "Validation of the Abstraction Index," A technique for testing and diagnosing copy flaws of any kind was described in Carter's article "The Content Response Code: A Pre-Testing Procedure."

Most of the studies above were conducted in universities and the procedures have been adopted by advertising researchers to varying degrees.

The testing of complete executions (whole ads) is one distinction between message research in the print media and in the broadcast media. In the print media, it is feasible to test individual elements of ads--separate tests for pictorial matter, body copy, headlines, etc. In the broadcast media, however, it is more often necessary to use complete executions of commercials since each element depends on and interacts with all the others. Lindley's article "A Behavioral Measure of TV Viewing" describes a pretesting method for TV commercials. Caffyn's article "Telpex Testing of TV Commercials" describes a quick and inexpensive method of testing with videotaped "rough" executions.

Pre-Testing, Post-Testing and Pilot Tests

One distinction that needs to be made in message research is that between pre-testing and post-testing (sometimes called pre-checking and post-checking).

Pre-testing simply means message research that occurs before an ad is published or broadcast in one of the media. It includes obtaining consumer reaction to: individual elements of an ad or commercial; a rough complete execution prior to production of the finished product, including TV "story-boards"; even a slick finished production before insertion in the media. Whatever form the pre-test ad takes, it is a way of estimating in advance, usually on a small sample of product prospects, how good it is likely to be. The cost of constructing an advertisement is generally much less than the cost of publishing or broadcasting it in a multi-million audience medium, particularly when it is planned to run the ad several times. It may be considered as cheap insurance against an expensive flop, even when several versions must be tested to come up with one final advertisement.

Post-testing simply means checking on the performance of an ad or campaign after it has appeared in one or more media. Surveys can determine how well an ad reached and influenced its audience in real life. While post-test results provide useful hindsight on ad performance, it has a couple of important limitations. One is that post-tests of ads as they appear in the media will reflect not only the effects of the ad itself but also some effects created by the vehicle in which it appears; it's difficult to separate those effects to get a precise reading on the message's quality. The other important limitation is that, even if a precise measure of an ad's quality is obtained, it's too late to do anything about it--the message has already appeared and can't be corrected if it had no effect or negative effects. Even the most skillful post-mortem doesn't help the patient who died.

Somewhere between the labels of pre-testing and post-testing are those cases where an ad is actually published or broadcast just once and the results tested to determine if the ad should be continued in the future for several repetitions. "Pilot test" is the best term for such a procedure. "Split-run" tests among a small proportion of a magazine's subscribers can help determine if it should be placed for complete circulation. Placement of an ad for complete distribution in a small-circulation publication can help determine if it should be placed in one or more large-circulation publications. Running a TV commercial in a single market can give clues to its effectiveness before placing it on a national network; the same can be done in the case of newspaper advertisements. While the results of such pilot tests are more expensive than pre-tests on small captive audiences, they have the advantage of appearing in vivo--under real-life natural conditions of audience exposure and selection. And they have an obvious advantage over post-testing--the ad can be corrected or modified before extremely expensive large-scale distribution is attempted.

"What Do You Mean, 'Advertising Effectiveness'?"

In terms of the advertising process model, an effective advertisement is one that creates maximum perception, communication, and eventually desirable consumer behavior. These terms are easy to define theoretically but more difficult to define operationally--that is, in terms of something that can be measured objectively.

First of all, it must be assumed that there is no single meaning for advertising effectiveness. Different ads or campaigns are designed to reach different kinds of people and to achieve different kinds of purposes. By way of illustration, a single advertiser would use different kinds of ads (a) to reinforce loyalty and repeat-buying among people who already buy his brand, (b) to introduce his product to people in a geographic area where the product is unknown, and (c) to capture customers who use a competitive brand. The purpose of the advertising in each case is different, the advertising message itself should be different, and the measure or criterion of effectiveness should be different.

Here are two useful rules for advertising effectiveness:

1. The purpose or goal of advertising must be clearly stated before advertising effectiveness can be defined. For a single advertiser, the goals will be defined differently in different cases.

2. The correct measure of advertising effectiveness must be defined in terms of the intended purpose in any particular case.

It is incredible, but true, that many advertisers and agencies will embark on an expensive advertising program without clearly defining their goals. In other cases, a single goal will be set when in reality several are needed. In other cases, the wrong goal or goals will be formulated based on inadequate knowledge of the true state of affairs. Such aimless advertising is irresponsible if not downright unethical.

Other advertisers, after having defined some specific goals, never bother to measure the results.

In advertising, even after specific goals have been set there is still the problem of deciding which of several possible methods of measurement is best. There is disagreement even among experienced advertising researchers on the "best" way to measure effectiveness. As in many other fields of education, the best course of action here is the eclectic approach--that is, several methods of measurement will be described leaving the eventual choice up to the researcher himself, depending on the situation.

Reactions to Product and Reactions to Advertising

When there is a specific goal for advertising, it seems logical to define effectiveness as the extent to which a message achieves that goal.

For example, say that one purpose of the advertising for "Sparkle-dent" toothpaste is to increase public awareness of it. To increase awareness or salience means to make it a name that pops into a person's mind when he thinks of buying toothpaste, to make it a familiar and salient brand that is automatically considered when the time comes to buy. An appropriate "awareness" question that has been used on many different kinds of products is "When you think of toothpaste, what's the first brand that comes to mind?" That is a product awareness question. But one can also inquire into awareness of the advertising with a question such as "When you think of toothpaste advertising, which brand's advertising stands out in your mind?"

There is a crucial difference between questions designed to measure responses regarding a product and responses regarding the advertising for that product. Past research has shown that a consumer can have high awareness, recognition, recall, and attitudes regarding a product's advertising but not regarding the product itself, and vice versa. Similarly, an ad can be interesting, believable and comprehensible but fail to sell the product involved.

It seems sound common sense to conclude that, since the purpose of advertising is to market the product, then the success of the advertising should be measured in terms of the product and not the product's advertising. Whether the advertising itself is remembered, identified, or thought of favorably is really irrelevant to the main purpose. However, consumer responses to advertising can be useful for diagnosing the advertising in terms of making it do a better job of marketing the product. Consequently, many surveys will ask questions about both a product and its advertising. But the researcher should always keep in mind that the creation of favorable reactions to the product is the primary concern of the advertiser, the principal goal of advertising, and consequently the principal goal of message research.

Disagreement and Agreement Among Experts

It would be misleading to claim that advertising research experts are in complete agreement on the best methods of research for any particular measurement problem. As evidenced by the contents of this book, there are numerous measurement methods; all of them are in use, to varying degrees.

But advertising researchers come in a variety of different backgrounds--psychologists, sociologists, economists, business administrators, statisticians and others. All are firmly committed to the improvement of advertising effectiveness through scientific methods, but they see many roads as leading to the same goal. This may not seem so strange upon considering that, even in some of the "hard" sciences such as chemistry and physics, the experts are far from unified on theories and research methods. In the "softer" sciences dealing with human behavior, the scientific approach is more recent and there is consequently less unity on theories and methods.

The considerable discussion and controversy over research methods is revealed in a survey of research directors from the 50 leading U.S. advertising agencies. The American Association of Advertising Agencies (AAAA) formed a research subcommittee--consisting of highly respected advertising researchers--to design a study to get expert opinion on various methods of message testing. The conclusions below are based on the 40 research directors who replied to the mail questionnaire.

Primarily the study aimed at evaluating nine different advertising measures:

> Advertising-related measures of recognition, recall, comprehension, liking, believability, persuasiveness
>
> Product-related measures of attitudes and beliefs (toward the brand or company), buying preferences, and behavior.

All these terms are described elsewhere so they will not be defined here. They were evaluated both as measures of individual ads (print or broadcast) and of advertising campaigns. Further, the experts judged the different methods according to their value as measurements of overall effects of a finished advertising execution rather than for pre-tests of specific aspects of copy. Keep in mind that these evaluations don't pertain to measuring advertising exposure or perception--rather to communication and effects.

Validity for individual advertisements. Here is how the nine measures were evaluated in terms of the number of research experts giving a rating of "highest value" to the measure.

Measuring Method	% rating it of "highest value" for individual ads
Comprehension	80%
Behavior	63
Attitude	63
Recall	63
Buying preference	53
Believability	38
Recognition	15
Persuasiveness	13
Ad-liking	13

Thus, a majority of advertising research experts believed that measures of comprehension, behavior, attitude, recall and buying preference are quite valuable for evaluating individual ads, in that order; on none of the measures was there complete agreement. Note also that most of the measures labelled here as "ad-ratings" (believability, persuasiveness, liking, etc.) have little value in the eyes of these experts; however, another ad-rating measure--comprehension of the ad--was believed to have great value.

Validity for advertising campaigns. For some odd reason, the AAAA study asked for evaluation of only four of the above measures, plus an added starter "awareness," for measuring the effects of advertising campaigns. Here is how those five measures were evaluated.

Measuring Method	% rating it of highest value for campaigns
Attitude	75%
Awareness	59
Buying preference	59
Recall	40
Recognition	15

A majority agreed that attitude, awareness, and buying preference are of great value in measuring campaign effectiveness. Recall and recognition have little value, according to the experts.

Validity related to test conditions. Several other questions were asked about research methodology in ad testing.

Exposure of test advertising should be under natural conditions rather than artificial conditions, by a two-to-one margin of opinion.

There was a two-to-one preference for experimental conditions of measurement (i.e., using a control group) rather than for comparative descriptive measurements. In effect, they favored controlled experimental designs rather than after-only or before-and-after designs.

Should ad tests use rough or finished executions? About half said it doesn't matter, and the others were equally divided between those preferring roughs and those preferring finished executions.

Are the research methods of different value in testing print and broadcast advertisements? Eight of the nine individual ad measures were judged of equal value for both print and broadcast testing. However, the recognition measure has higher value in print ad testing according to one-third of the researchers, the remainder feeling it makes no difference.

Should, and can, sales be the final criterion of campaign effectiveness? Even on this vital issue the research experts were in disagreement--exactly half felt that sales is the ultimate measure, the other half did not. Furthermore, only about two-thirds of the experts believed that sales can be measured with sufficient precision for testing advertising; this may reflect the fact that sales are indeed easier to measure for certain products than for others.

Some additional comments in the committee report should put the above findings in better perspective.

> These conclusions are based on consensus, but consensus is not in itself a guarantee of the validity of any measure. . . . There appears to be a dilemma on some measures between the better measure and the more practical measure . . . some measurement methods and techniques are much more a compromise rather than superior.

It was made clear by the experts that individual advertisement testing is done for a different purpose than campaign measurement. Copy testing of individual ads is performed as a diagnostic procedure, to determine "why" an ad does well or poorly and to provide remedial information. It is unreasonable to expect a single advertisement to produce sales, expect in special cases.

In any event, the lack of complete agreement among the experts indicates why so many different ways of measuring ad effectiveness are used, many studies using the "grab-bag" approach of measuring everything possible. Further methodological research will some day clarify some of these problems.

CHAPTER IX. PRE-TESTING MESSAGES

Pre-testing simply means testing an advertisement, or some part of it, before it is published or broadcast. One can pre-test selling concepts, themes or headlines, graphic presentation techniques, persuasion strategies, body copy or whole executions. One can obtain different kinds of responses--verbal, physiological or behavioral. One can measure reactions to ads themselves or to the products contained in the ads. One can present the test materials in homes, testing studios or laboratories, or to individuals and groups in their natural habitat of stores, airline terminals or waiting rooms.

In short, there is practically no end to the different ways in which ads can be pre-tested. This section contains a discussion of some of the most important kinds of pre-testing situations. A thoughtful reader can fill in with other likely variations and combinations of pre-testing procedures.

Pre-testing is usually concerned with testing the relative performance of alternative executions compared with each other, since there are few if any absolute measures of message effectiveness.

General Pre-test Guidelines

While there is much discussion and little agreement among advertising researchers on specific techniques of pre-testing, all will agree that pre-testing is desirable. (Curiously enough, a great deal of advertising is not pre-tested.) And most would agree with these general guidelines for copy testing, as expressed by large-agency research directors at recent conferences.

1. "Copy research is meaningless and even misleading unless it is conducted among the target prospects for the advertised brand." Several studies have shown that the responses of the general population, or non-prospects, are different from prospects.
2. "Copy research is inefficient and sometimes quite ambiguous unless preceded by adequate information about what to communicate and to whom." An integrated research program is viewed as consisting of: market and motivation research, to define the goals for a particular product, at a particular stage in its life cycle, in its particular competitive environment; concept research follows, to determine the best message in light of these goals; copy research deals with the particular creative execution of the basic concept; effectiveness measurement is needed to track advertising performance over time in the total market environment.
3. "Copy research is incomplete if it measures only responses to the advertising; measurement must also include responses to the product." The attention value of the advertising itself as in real-life exposure, plus the effects in terms of response to the product, are crucial to successful measurement.
4. "Copy research is sterile unless it contains information on the elements accounting for the responses obtained." It's not enough to know which ads work and which don't; diagnostic information on why they were successful or not is essential to true understanding.
5. "Copy research, if it is to be predictive, must validate its methods." A copy test or method is misleading and even harmful

if the results in real-life are different from those obtained in the test. Somewhere along the line, validity of copy testing methods should be demonstrated through basic research, not merely assumed.

6. "Most copy research deals with single ads or commercials. Ways must be found to pre-test campaigns." Because of the minimal effects of single messages in the real world, single-message tests may not reveal much.

7. "The copy tester should work closely with the business and creative groups of the advertising agency." This makes him more clearly aware of wider range of alternative ideas for testing, and increases the actual use of the research results in creation of the advertising.

8. "Copy performance measures must be found which reflect the sales goal of advertisers." The ultimate goal of advertising is sales, and--even though advertising is only one of the forces influencing sales--methods of pre-testing ads and campaigns should use measures that are related to sales; this requires research which measures sales and communications criteria simultaneously and identifies those communications measures which are most relevant.

9. "Copy research should lead to the development of general principles of ad construction." Research to satisfy curiosity regarding immediate problems and individual ads must be standardized so that a systematic body of knowledge can be built up, aided by thorough record-keeping on ad performance.

For the person who wants to go beyond being simply a technician, those are the kinds of problems which can provide substantial challenges and rewards.

The importance of concept testing. Many researchers maintain that adequate pre-testing of selling concepts is more valuable than testing of themes, graphics, persuasion strategies, or whole ad executions. For one thing, this approach provides information at the earliest possible stage in development of an advertising campaign, before the considerable expense of creating rough or finished advertisements, either print or broadcast. Many different ideas for selling a product are always generated at the beginning of an advertising effort--there is much evidence to suggest that "what you say" is more important than "how you say it."

Concept pre-test procedures have been used for several purposes--for selecting the best concepts out of many available, for stimulation in creating new concepts, for diagnostic information to help the creative staff make more effective translations into finished copy.

In most concept testing techniques, the prospect-respondent is exposed to one or more statements, each focussing on a single idea, usually without mention of the specific brand but concerned only with the product itself. Verbal responses from personal interviews are the usual thing, but physiological measurement in other research situations is also used.

The most common methods of presenting the concepts are: paired-comparison, rank-order, guided-usage and own-brand comparison.

In the paired-comparison procedure, a respondent is presented with two different product concepts, told that they represent different brands of the same product, then asked to express his preference and the reasons why.

In the rank-order method, a respondent is presented with a series of product concepts, asked to rank them in order of preference and the reasons for his choices; this is different only in degree from the previous method.

The guided-usage test is quite different. The respondent is exposed to a product or given a sample of it for a trial usage, along with a single concept. After trial he is asked to comment. Different concepts are evaluated by using matched samples.

In the own-brand comparison method, the respondent is asked to evaluate one or more product concepts against the brand he now uses.

Other researchers maintain that selling concepts can only be tested in the form of complete ad executions. Regardless of the method of testing, however--many more of which will be discussed in the sections that follow--the advertising researcher should be aware that concept testing may be the most important stage of investigation.

General Pre-Testing Methods

The group discussion method. Small groups of respondents are sometimes assembled for a round-table discussion of an ad or a series of ads. While this is not presumed to be a typical advertising exposure situation, the interaction and stimulation of several consumers on each other is sometimes helpful in diagnosing weak and strong points of an advertisement. The primary benefit from such discussions is likely to be in what it reveals about consumer motivations and only secondarily an evaluation about advertisements used as stimuli.

This method--also known as the "consumer jury" or "focuses group" technique--is most often used in the initial stages of planning an ad or campaign. It provides a creative stimulus for the artists and writers responsible for producing an ad, and it simultaneously keeps their ideas grounded in the reality of how product users think and feel. This latter point is very important, since advertising people usually live and work in a world which is quite different from that of the typical consumer.

While group members are sometimes asked to give numerical ratings or rankings of ads used in the discussion, these are not reliable as quantitative indices for a variety of reasons--too few people in the group for statistical significance, undue influence of some individuals on others, the abnormal exposure situation, and others.

Projective methods. Another way of getting at the deep-down motivations, desires and preferences of consumers is to get their reaction to vague and ambiguous stimuli that can be interpreted in a variety of ways. Thus, instead of showing an advertisement, the researcher might show neutral fragmentary pictures, words, or situations and ask the respondent to tell a story about it, fill in the details, complete it. In short, the reasoning is that the respondent will "project" himself into the scene and reveal his underlying thoughts or feelings.

The principal types of projective methods are association techniques, construction techniques and completion techniques.

In the association techniques, the researcher presents a word, an ink-blot (the Rorschach test) or some unidentified shape such as a cloud, and the consumer is instructed to respond with the first word, thought or image that occurs to him. An illustration of word association: "What's the first thing you think of when I say HOT? FLOOR? AUTOMOBILE? HEADACHE? SPARKLE-DENT TOOTHPASTE?" (and so on).

In the construction techniques, the researcher presents a scene or picture, and the consumer is instructed to make up a story about it; the Thematic Apperception Test and the Blacky Pictures are standardized versions of this technique. An illustration: "Here is a picture (picture shows woman handing tube of toothpaste to man about to brush his teeth). What are they saying to each other?"

In the completion technique, the researcher shows an incomplete sentence, an incomplete picture or some other unfinished stimulus-object and the respondent is instructed to fill in the details or complete it in any way that he wishes. An illustration of sentence-completion: "The ideal toothpaste is ___________. Most toothpastes are ___________. Sparkle-Dent toothpaste is ___________."

These are only a few of the different kinds of projective methods, variations of which may be called "unstructured" methods, "depth" interviews, or "motivation research" methods. They are not so often used in testing complete ads as in getting ideas for selling concepts or persuasive copy.

The crucial points in the use of projective methods lie in (a) selecting the proper stimulus to present to the respondent, so that he will talk about the topic desired, (b) making the stimulus neutral or ambiguous so that the respondent projects himself into the situation rather than simply describes something already present and (c) interpreting the respondent's remarks in a meaningful and objective manner so that powerful advertisements can be constructed from them. There has been much disagreement about the value of motivational research among researchers on this latter point--that is, if different researchers interpret the same respondent's remarks in different ways, one can seriously question the validity of the results.

However, questionable projective techniques may be as research tools, they are undoubtedly helpful as creative tools--that is, as sources of ideas for the people who produce advertising, or for researchers as a basis for formulating structured tests and measures.

Laboratory apparatus methods. Specially designed laboratory equipment can obtain non-verbal or physiological responses to ads in any stage of development. The equipment is usually designed both to present an advertisement under strictly controlled conditions and to record simultaneously the reaction of the respondent.

The eye-movement camera is one such device used primarily for print ads. A respondent's head is fixed into a particular position so that, as an advertisement is presented, a special camera can photograph his pattern of eye movements about the ad--the place of first fixation, second, and so on, so that the eye-movement pattern can be charted just like car tracks over soft ground. By testing a number of respondents in this way, the researcher can deduce which parts of the ad are being concentrated upon or ignored, the best sequence for organizing the ad's visual elements, and so on. Having found the "weak" spots in an ad, the ad may be redesigned then re-tested until the desired effects are achieved.

The galvanometer or psychogalvanometer is a device attached to a respondent's hand to measure changes in electrical resistance. When shown something exciting, a respondent's nerve activity will result in perspiration which in turn increases electrical conductivity, just as any object becomes a better electrical conductor when standing in a puddle of water. When a respondent is relaxed, there is virtually no sweat. A respondent viewing several different ads will involuntarily reveal variations in interest through automatic recording of his electrodermal response (EDR). One problem, however, is that both an extremely unfavorable reaction and an extremely favorable reaction will cause excitement and show up in the same way on the graph.

The electroencephalograph is a device which measures the electrical output of the brain, and changes are revealed by fluctuations on a graph. Electrodes are attached to the head, then as various ads are presented a graphic record is made of the "brain waves" which result.

The eye pupil camera is one of the more promising devices used in testing print advertisements. A specially designed camera measures tiny dilations and contractions of pupil diameter. Pupillary dilation, or increase, indicates a favorable response while contraction indicates a negative response.

Some other laboratory measures of physiological response to ads include: respiration (breathing rate), pulse rate, blood pressure, glandular biochemical changes (blood sugar, adrenalin, etc.), uterine contraction, salivation ("drooling") and muscular contraction.

These laboratory devices to measure physiological response are not always precise enough to measure the rather small changes usually evoked by advertisement. And, as noted, there are often problems in interpreting the data. However, the eye movement camera and the eye pupil camera have become important tools in pre-testing; with refinements in equipment and technique they may become even more important.

The tachistoscope is a device for presenting print ads for any desired duration--one-thousandth of a second, one-hundredth, one-tenth, and so on. The rationale is that the quicker an ad can be perceived and comprehended, the more attention-getting it is likely to be in a real-life advertising situation, since readers are known to scan through a magazine or newspaper rather rapidly. Thus, a group of respondents may be presented with several ads, and everything else being equal, the one that is recognized fastest is probably superior in attention value. The tachistoscope starts by presenting an ad so quickly that no one can see it, then the duration time is gradually increased until a respondent arrives at his recognition threshold. The "communiscope" is one model of the tachistoscope specifically designed for advertising tests.

The "program analyzer" is a laboratory device which measures conscious reaction rather than involuntary physiological response. When exposed to a broadcast commercial or program, a respondent indicates amount of interest by twisting a dial (clockwise if enjoyed, counterclockwise if not), or by moving a switch to left or right, or by pressing one of a series of buttons. An auditorium or studio can be fitted with several dozen or hundreds of response stations for group testing. A continuous graph of group reaction can be automatically and continuously recorded, indicating the affect (i.e., degree of liking or disliking) for any portion of the broadcast stimulus. At points where the graph line plunges down, changes can be made to improve the ad or program performance; verbal questioning of respondents on their reasons for liking or disliking certain portions provides valuable diagnostic data for improving effectiveness.

Binocular devices. "Binocular rivalry" is the term used for presenting two ads simultaneously to determine relative attention value. When extraneous factors such as eye dominance, illumination and distance are controlled, the relative attention value of several ads can be determined by presenting all combinations of pairs through the binocular device. In each pair, the respondent will perceive only the dominant ad. This apparatus has been used primarily in testing alternative graphics techniques in print ads--layout, color, pictorial elements, etc.

These laboratory devices may be used either singly or in combination. For example, one commercial testing firm--Audience Surveys, Inc.--uses the program analyzer and the galvanometer in combination with verbal questioning to pre-test TV programs and advertisements.

As in other forms of research, the validity of the data depends on the extent to which the pre-test respondents are representative of the larger population for which the advertising is intended.

Other general research methods. Many other variations of general research methodology are adaptable for pre-testing. At this point, it may be useful to review the earlier chapter on "The Most Common Advertising Research Plans." Sample surveys, laboratory experiments, controlled observation, content analysis, controlled field experiments, and attitude measurement--all have their uses in pre-testing.

Broadcast Message Pre-Testing

In addition to some of the general pre-testing methods already described, several other methods are available for testing broadcast commercials and programs. For some odd reason, very little pre-testing of radio materials has been reported, hence this section will concentrate on television; however, most of these methods would be equally well-suited for radio.

Commercial form for testing. One important question for advertising researchers in this connection is: Must a TV commercial be in slick finished form for testing, or will a preliminary rough version do the job? The question is important because the production of a finished commercial can be very time-consuming and expensive--therefore, the use of rough preliminary versions would save much time and money.

Three principal forms of commercials for testing are most common--storyboard, rough, and finished.

The storyboard form simply shows in sequence a number of still scenes which convey in sequence the "plot," the principal images and words involved in the planned commercial. The scenes are sometimes artist's renditions, sometimes photographs. A projector may be used to flash the stills on a screen in rapid sequence--as in film-strip showings. Sound may be superimposed over the visual scenes--music, narration, sound effects--to further simulate a real commercial. There is a wide range of polish possible with storyboards, ranging from simple sketches of ideas in the earliest stage of development to painstaking executions including all commercial elements.

The rough commercial form for pre-tests is similar to the commercials actually seen on television. However, instead of highly-paid top-flight announcers and actors, they may use whoever happens to be handy; instead of expensive filming "on location," studio scenery or a stock film background may be substituted. The script may not be as finely polished as in the final version. In short, it is a preliminary "rough draft" of the final product; several rough commercials can be produced within the cost and time required for filming a single finished commercial. In this way, the essential ingredients of several ideas can be tested, the obviously ineffective ones thrown out and the promising ones modified, to increase the selling power of the final version.

Sometimes advertising agencies produce their own rough commercials, sometimes they hire outside firms to do the job. One firm that specializes in both production and pre-testing of rough commercials is Telpex, Inc. Telpex's "instant production" facilities make it possible to video-tape several commercial roughs in a single day, view the commercials immediately after shooting, make changes in them as desired, and then re-shoot them with improvements. These can be shown immediately in a test theatre next door to specially selected audiences of about 50 persons at a time. Audience reaction questionnaires can be analyzed quickly and reports produced soon thereafter.

This procedure eliminates one of the major problems in pre-testing TV commercials--time. In the space of a few hours or days, several ideas can be produced, tested, and a decision reached for final production.

A finished commercial can also be used for testing, of course, and some researchers argue that this is the only way to get an accurate evaluation. It probably does provide a somewhat more valid measure of effectiveness than less-finished versions, but the question is "How much more valid?" and "Is it worth the extra cost and time?" Then too, there is a psychological hazard in using finished commercials in testing since agencies and advertisers are reluctant to change or throw on the scrap heap a finished commercial, even if test results show it to be ineffective.

In practical terms--considering production cost, media costs, and the low probability of success for untested commercials--the pre-testing of several storyboards or rough commercials to produce one effective finished commercial would appear to be a wise investment.

Different TV pre-testing environments. Another question that arises in pre-testing TV commercials concerns the situation or environment in which the respondent views and responds to the test message.

The most common test environments are auditorium tests, in-home tests, on-the-air tests, and split-cable tests. They may be used for testing complete program formats or commercials or both.

In auditorium testing, a group of people may be deliberately assembled in a studio, theatre, classroom or auditorium, or portable trailers with a few seats and projection equipment may be moved around to catch consumers in shopping centers and other natural habitats. This "captive audience" technique is the most frequently used test environment. It lends itself to obtaining various kinds of verbal or physiological responses from a number of people simultaneously.

The commercials to be tested are usually buried in other content to disguise as much as possible the real intent of the procedure. For example, the audience may be shown a real movie or TV program instalment with several different commercials interspersed, though only one is the object of testing. In addition to tests of different commercials, the auditorium method is used in testing different presentation techniques--e.g., placement of commercials in programs, length of commercials, and so on--by having a different group see each alternative.

The researcher may gather ad-rating information (enjoyment, interest, persuasiveness, believability, etc.), recall of sales points and other verbal responses. The Schwerin Research Corporation--formerly one of the leading testers of television commercials--disguised the purpose of the assembly as much as possible, and obtained brand-preference ratings before and after exposure to the test message; the change in preference was considered to be a measure of the selling power of the commercial.

Another testing firm--Audience Surveys, Inc.--uses laboratory equipment, such as the program analyzer and the galvanometer to obtain diagnostic data in auditorium tests. These responses may be combined with ad-rating responses, brand preference change, and other verbal responses to get a comprehensive evaluation and diagnosis of different commercials.

The in-home test utilizes a portable lightweight film projector that can be carried by an interviewer to individual consumer homes. The test film is shown and the respondent is questioned as in some of the other survey situations.

The on-the-air test involves actual broadcast of a test commercial in a regular TV program format in one or more cities. The local station cuts in on the network program at commercial time and shows the test commercial instead of the one being nationally or regionally televised. Sometime later, personal or telephone interviews are conducted in a sample of area homes.

The split-cable method of presentation is considered by many researchers to be the most valid method of measuring the effects of a commercial. Within certain cities, it is sometimes possible to cut in on a broadcast and simultaneously beam Commercial A to one group of homes and Commercial B to another matched group. Or one group of homes can be "blacked out" and receive no commercial at all, and thus serve as a control group. Interviews are then conducted with a sample of respondents from both groups of homes.

Each of these different TV pre-test environments has its advantages and disadvantages. In auditorium testing, the primary advantage is in the tester's complete control over all details of commercial transmission and reception, and in the feasibility of laboratory-type response equipment; primary disadvantages are the unnatural "forced exposure" to the commercial and the non-typical sample of respondents (people who will appear for such a test are different in many respects from people who refuse). The in-home test can come up with a representative sample of respondents, but the viewing situation is quite artificial. The on-the-air test shows the commercial under natural viewing conditions, but the lack of an adequate control group makes it impossible to isolate the effects of the test commercial from the many other influences in operation. The split-cable method comes closest to attaining representative samples, natural viewing situations, and control over extreneous influences--however, the cities in which it can be used at present are few and probably not typical.

Many major advertisers have conducted exhaustive investigations into the various methods of pre-testing TV commercials. All have reluctantly decided that the perfect method is yet to be found. Consequently, some use one method and some another.

Print Message Pre-Testing

The "ad-rating" method. Present the respondent with several print ads, either simultaneously or in sequence. He is then asked to rank the ads (first, second, third, etc.) or rate them (on a 100-point rating scale, for example) on such dimensions as reading interest, enjoyment, believability, attractiveness, persuasiveness, good taste, useful information, stimulation of desire for the product, and so on. Or he may simply be asked to say which he likes best, which he likes least, and why.

A variation of this procedure is the "paired-comparisons" method. The respondent is first shown X and Y together and asked to choose one that is most enjoyable, persuasive, etc.; he is then shown the XZ pair, then the YZ pair, and repeats the rating process for each pair. Some researchers maintain that people can't reliably judge or discriminate among more than two objects at the time; therefore when getting judgments on three or more ads, they recommend showing the ads to respondents in pairs.

The "ad-rating" method is also sometimes known as the "order-of-merit" method, because it basically requires a respondent to rank a number of ads in order from best to worst along some dimension of judgment. This method is frequently used for "quick and dirty" tests of alternative advertisements. But it is of dubious validity, for all the reasons mentioned in a preceding section on "How Not to Determine Ad Effectiveness."

The portfolio method. This is a method of testing complete ads in which a test advertisement is presented to respondents along with several others in a folder, or portfolio. The respondent is asked to look through the portfolio, taking as much time as he likes, and then is questioned about which ads he recalls, which products were advertised, what product facts he remembers and so on.

The crucial difference between this and the "ad-rating" method is that the respondent isn't aware which ad or ads are the real subject of inquiry, and the response is considered more similar to what would occur in real life.

In the case of three test ads--X, Y, and Z--three portfolios may be made up, alike in every respect except for the test ad. A group of respondents is randomly divided into three samples, then each sample sees a different portfolio. Since the portfolios are alike except for the one test ad in each, the difference between the responses to X, Y, and Z indicates the differential effectiveness; responses to the other ads (called control ads, since they are the same for all groups) should be equal for all three groups.

The portfolio method is one way of disguising the real purpose of the researcher and therefore is supposed to result in more valid measurement than if the respondent knows what's expected of him.

Another version of the portfolio method is the "dummy magazine." A test ad is bound into what appears to be a real magazine--complete with cover, articles, stories and other ads. Some commercial testing firms regularly publish a dummy magazine for the specific purpose of testing ads. This is considered a better research disguise than the ad portfolio.

The split-run method. This is a way of testing two or more finished executions of print ads in a controlled field experiment. It might more properly be called a "pilot test" rather than a pre-test. In its simplest form, a publication's printing press can be set up to insert Ad A in one-half the copies, and Ad B in the other half. After publication and circulation, interviews can be conducted with the two groups of respondents and the differential responses to the ads compared. The split-run can sometimes be arranged to alternate ads A and B to every other name on a subscription list. Somewhat easier to implement is the "every other market" technique, in which all copies going into one group of cities get Ad A and all copies going into another matched group get Ad B. Then there is the "checkerboard" design--usually used by newspapers--in which a single city is divided into squares or segments like a checkerboard, and each segment gets a different ad version. The *Milwaukee Journal* has pioneered in the development of this form of ad testing and set up a complete production, distribution and research procedure to accommodate advertisers wishing to test their ads.

Instead of conducting interviews to measure response to the ads, *coupon returns* are frequently used. A coupon, printed as part of an ad, is coded so that it's easy to determine the number of returns produced by each of the test ads. This is a particularly effective method of measurement for several reasons. It doesn't rely on verbal reports, and little or no research expense is incurred. The coupon return is an undeniably valid measure of effectiveness, particularly for businesses which rely heavily on mail-order sales. International Correspondence Schools has a continuous ad testing program using coupon returns from split-run experiments. Over the years, ICS has accumulated a considerable knowledge of the principles of good advertising for their purpose, based on such research results.

Coupon returns are also known as "inquiries." Sometimes prizes or premiums or free samples are used to increase inquiries. The validity of inquiries as measures of ad effectiveness depends on the extent to which the inquiry is related to sale of the product. In the case of a mail-order business, the relationship between inquiry and sales is direct and there is no doubt of the validity of the measure. However, free premiums or contests may attract large numbers of inquiries and not have any effect on product sales; thus, in the case of non-mail order businesses, additional research may be necessary to determine whether coupon returns and inquiries have any relationship to increased sales of the product.

Content response code. This print ad pre-test method is a combination of content analysis and consumer survey. It has been used primarily for written material, though it has also been used for ads combining written and pictorial matter. It operates like this: a small sample of the intended target audience is selected; each one is instructed to read the test ad in its entirety and to mark with a "minus" sign any segment that they find disagreeable, hard to understand or negative in any other way; similarly, he is to mark with a "plus" sign those segments that are highly agreeable, persuasive or interesting any any way. The combined judgments of the respondents are then used to plot a graph of the whole message--quite similar to the "program analyzer" graph produced by mechanical equipment in pre-testing broadcast programs and commercials. Where there are serious "valleys" or weak spots in the ad, the respondents are questioned further on their reasons for judging it negatively. Corrective changes are made, then the whole ad is re-tested with another sample of the target audience.

This method was developed by a university researcher, Dr. Roy Carter, for use with U.S. Government propaganda leaflets. Though it has not been widely used in advertising, enough has been done to indicate that it has considerable potential for improving print ad effectiveness.

Direct mail method. This is a slightly different form of "split-run" using mailed ads to lists of prospects or other special target groups. Two or more alternative ads can be pre-tested for pulling power by randomly dividing a mailing list into two or more sub-groups as needed. Differential effectiveness may be measured by the number of mail orders returned, by questionnaire surveys of the recipients, or by coded "money-off" coupons to be turned in by recipients when making a personal purchase. An additional possibility with direct-mail is that different advertising appeals can be tested among samples of the same group, or the same appeal can be tested among many different kinds of groups; specialized mailing lists for almost any known kind of person can be purchased--barbers, teenagers, sports-car owners, or whoever.

One assumption in the use of direct mail, as a pre-test for magazine and newspaper ads, is that the results would turn out the same if actually published in the desired medium; this assumption may or may not be correct unless validated by additional research. Once validated, however, direct-mail can provide an extremely accurate indication of differences in ad effectiveness. It is as adaptable for small-scale local advertisers as for large national concerns.

"Advertising Puzzle Game". A print ad layout is developed, but with a blank space where the main illustration is to appear. The incomplete ad and several different illustrations are presented to a respondent, who is then asked to complete the puzzle with the most appropriate illustration. The same procedure can be followed with headlines, or other ad elements.

As mentioned earlier, the number of pre-testing methods is limitless, or limited only by the creativity and imagination of the researcher himself--and most competent researchers are quite creative, adapting basic existing methods to fit particular situations. The preceding discussion has covered most of the common procedures.

Outdoor Message Pre-Testing

Many of the pre-testing methods previously described can be adapted for testing posters and car-cards. Group discussions, projective techniques laboratory apparatus, content analysis, controlled observation, auditorium testing, ad-ratings, and split-runs--all have outdoor applications. Since, in the unique case of outdoor, the medium and the message are identical, some message research devices have already been discussed in the section on outdoor media research.

Since outdoor ads must capture attention and convey a message more quickly than most other mass media, the most useful research methods may be found among the laboratory apparatus methods designed to measure visual responses--the tachistoscope, ocular devices, eye-pupil camera, and so forth.

The "simulated film trip" is one particularly ingenious instance of outdoor pre-testing. In one privately conducted study, the procedure went as follows: a well-travelled highway route into a city was fitted with several test posters on various billboards along the way; a movie camera was mounted on an automobile which then drove the route just as any incoming vehicle would, and a film was made showing the trip as it would be seen by the driver. In the test situation, respondents were shown the film--under pretext of some other purpose--then questioned regarding what they remembered about the trip. Specific questions were asked regarding interesting posters along the way, brands advertised, and so on. This method is adaptable for in-home testing, auditorium testing, and group discussion.

CHAPTER X. POST-TESTING MESSAGES

Post-testing has been defined as checking on the performance of an ad or campaign *after* it has appeared in one or more media. Its primary purpose is to provide hindsight on message performance and to draw some conclusions regarding the conduct of future advertising. As previously observed, post-testing results are usually a mixture of media effects and message effects, and interpretation of such data must take this into account.

In terms of the advertising process model, post-testing deals with advertising exposure, perception, communication and occasionally with sales results. However, it most commonly deals with the advertising exposure and perception stages: Was the audience member exposed to the ad? How much attention was paid to it? What was remembered from that exposure?

The additional stages of communication and behavior can only be measured indirectly through experiments to determine if exposure and perception resulted in increased brand awareness, improved brand attitudes, more brand sales.

To summarize, post-testing can be separated into two roughly distinct categories of measurement:

1. *Direct measurement* of advertisement exposure and perception, usually determined by measures based on memory, such as recognition, aided recall, and unaided recall of the *advertisement*.
2. *Indirect measurement* of communication and behavior resulting from the advertisement, usually determined through changes in awareness, attitudes and sales behavior regarding the *product itself*.

Recognition and Recall of Print Ads

Direct measurement of advertisement exposure is quite similar to the vehicle exposure studies discussed under media research; in fact, vehicle and advertisement exposure are usually ascertained in the same survey interview. Once exposure to a vehicle has been established, the interview may use recognition questions or recall questions to determine ad exposure.

Recognition and the Starch Score. In studies of print publications, a respondent leafs through the pages and points to the items he remembers noting or seeing, and tells an interviewer how much of it he read. The results for a sample of persons are tabulated according to: percent who saw or noted; percent who read part; percent who read most or read all. These are sometimes called "reader traffic scores" and are used to evaluate both advertising and non-advertising (editorial) items.

Daniel Starch and Company has developed a standard service for researching reader traffic scores, and every advertising man knows what is meant by a "Starch Score." Starch scores are basically the same as those described above, except they are reported as "Noted" (the percent of publication readers who saw the item), "Seen/Associated," the percent who read any part of the ad containing the brand name; "Read Most" (the percent who read more than half of the ad's written copy).

The Starch service also includes an annual summary of reader traffic scores called the Adnorms report. This provides average reader traffic scores reported separately for each magazine by product class, size of ad, and use of color. An individual ad can be evaluated by comparing its score against the norm or average for all similar ads (same magazine, same size, same use of color, same product class).

Each Starch survey consists of interviews with 200 to 400 persons--half men and half women--from 20 or 30 areas of the United States. In addition to a printed report on issue readership scores, Starch prepares an actual copy of the magazine issue surveyed, with stickers on each item showing the "noted," "seen/associated" and "read most" scores with separate stickers for each portion of the ad.

Controlling recognition errors. For many apparent reasons, recognition measurements as described above are subject to reporting errors by the respondent. Some of the reasons for error are: genuine confusion with other advertisements; confusion as to which magazine or which issue actually contained an ad; guessing; deliberate exaggeration; eagerness to please the interviewer; misunderstanding; interviewer errors; and so on.

Two methods have been devised to arrive at more reliable recognition scores; both require additional research of an extremely careful nature. The method of prepublication control requires that a recognition survey of previously unpublished ads be conducted before the ads actually appear in print; these ads may be presented in a portfolio along with already-published ads, in a dummy magazine, or in an old copy of the actual magazine. These "false recognition" scores for the unpublished ads are later subtracted from the scores obtained in the real recognition survey conducted after the ads are published. The confusion control method consists of having some unpublished ads surveyed along with the real ads--usually in portfolio--and getting recognition claims as usual; however, those respondents who consistently claim recognition of the false ads are later dropped from the analysis of the real ads.

In small-sample laboratory studies, a couple of other methods have been used to increase the accuracy of recognition claims. The tachistoscopic method is to show at high speed either complete ads or ad portions and require that respondents furnish further ad information before being certified as readers. The screening method consists in having several screens over an ad, then removing one at the time and getting recognition claims at several levels of visibility.

There are many variations on methods for refining claimed recognition scores. However, it is doubtful that the regular recognition measurement figures create any great problem as long as the principal objective is the relative evaluation of different ads in the same issue; the number of over-claimers is not greatly different than the under-claimers, overall.

Aided recall and the Gallup-Robinson score. Those who complain about error possibilities in the recognition method can get at least theoretical comfort from the more complicated aided recall procedure. Just as in the other method, respondents must first qualify as readers of a particular issue; the basic difference is that the test issue is kept closed and the reader must rely almost entirely on memory for his reports on readership of individual ads. Gallup-Robinson is a commercial research firm with a standardized service to provide such figures.

The aid for recall comes in the form of advertisers' or brand names; a list of such names is presented to the respondent with the comment that some advertised in the issue and some didn't. He is then asked to specify which ones he remembers from that issue, what the ad looked like, what it said, and so on. Responses are recorded verbatim. The fact of recall is not decided until the interview forms are edited back in the home office. The editing is a tedious and error-prone procedure which establishes whether or not an ad was actually seen, the most memorable parts of the ad, and a special score called Proved Name Registration (PNR). PNR represents the number of verified readers of an ad who recall some crucial element of the ad; this raw figure is adjusted for ad size, color, placement in issue, and number of other ads in issue. The adjustment, it is claimed, transforms the raw score into an absolute measurement so that ads from different issues may be compared by the same yardstick.

Because of the contrariness of human memory, however, aided recall scores are generally considered to be underestimates of true attention and readership.

Comparison of recognition and aided recall scores. Since readership of printed ads is, and has been, a subject of great interest to the advertising industry, some discussion should be given here to a basic methodology study conducted by the Advertising Research Foundation.

Discrepancies between recognition scores and aided recall scores on the same ads gave rise to the PARM study--Printed Advertising Rating Methods. The importance of the study stemmed from a fundamental concept; an ad cannot sell unless it is read. ARF members voted for PARM as the most-needed project at that time, the late 50's. A single issue of Life magazine was selected for an intensive study of both methods. One discrepancy found in the two methods was in the number of issue readers. The aided recall (Gallup-Robinson) method results in one-third fewer readers of a magazine issue--i.e., vehicle exposures--than the recognition (Starch) method.

Another discrepancy was in the characteristics of issue readers. The recognition method produced lower proportions of persons who: are under 40 years old, are in the middle and upper economic brackets, have college educations, are primary household readers, read their magazines at home, and spend only a short period of time with the issue.

In short, the two methods picked up different numbers of issue readers and different kinds of readers. Since recall had more exacting requirements (not necessarily more accurate), it in effect eliminated many persons of certain characteristics from the audience.

The major difference, however, was in the average ad readership scores obtained by the two methods. Recognition gave an average ad-noted score which was six times the average PNR score--19.2% to 3.0%. Furthermore, the methods produced different kinds of ad readers: recall ad readers tended to be younger, have lower income, lower education, lower occupational status.

With the passage of time between reading and the readership interview, recall scores dropped steadily while recognition scores held steady.

The basic conclusions of the PARM study are that recall gives much lower ratings, which are much more sensitive to three important methodological factors--the lapse of time before interviewing, the competence of interviewers, and the type of sample. This results in a lack of reliability, or "test-retest" consistency, in recall scores. That is, if the same ad is tested twice by aided recall, different scores are more likely to be obtained than through the recognition method. This was later verified in another study by a major advertiser. In short, differences in aided recall scores of different ads are more likely to be due to unstable research methods than to differences in the ads themselves; therefore, recognition measurement is the better technique.

Recall of Broadcast Ads

Recognition measurement of broadcast advertisements is not feasible for large samples of persons, since it is difficult if not impossible to show in an interview all programs broadcast even for a single evening. Consequently, recognition measures are not used in broadcast post-testing.

Telephone survey methods. However, the relative ease of recall tests of television advertising has resulted in widespread use of the method. The interviewing can be done quickly, usually by telephone, and is less taxing on respondent time than the corresponding magazine procedure.

A random sample of telephone names is drawn in a community. Within two hours after a program is completed, telephone interviews are conducted along with following lines:

> "Did you watch television tonight?"
> "Which programs did you see?"
> "Tell me something about (program seen)"

After being verified as a program viewer in these questions, exposure to the commercial is determined.

> "What products do you remember having seen advertised?" (unaided)
> "Here is a list of different kinds of products: Do you remember seeing commercials for any of them?" (aided)
> "What did the commercial tell you about (name of product)?"
> "What brands were advertised?"

Additional questions may be asked about the believability, importance, persuasiveness, etc., of the commercials viewed.

Through this procedure, the researcher can establish commercial exposure scores expressed in percentage terms; these may be interpreted in the same manner as magazine ad readership scores.

Diary method. This is not usually considered as a recall method, however it qualifies technically. In the diary method previously described under media research, the respondent may be asked to keep records of the commercials seen along with program viewing.

Conclusions about Recognition and Recall Post-Testing

Post-testing based on recognition, aided recall and unaided recall are tests of what people remember about specific advertisements.

The recognition method appears to be superior to recall as a measure of exposure and attention to ads; however, it tells nothing about the effects of the ad; for example, an ad can get high attention and readership, yet have no effect on attitudes, buying, etc.

The recall method, aided or unaided, is probably unreliable as a measure of exposure and attention; theoretically, it is a measure of an ad's effects on knowledge. However, research methodology is not precise enough to place great reliance on the validity of the results. Furthermore, even if the methods were precise, there is much evidence that factual recall of an ad is not at all related to the desired effects--that is, high factual recall of an ad may be accompanied by no change in attitudes or sales.

In short, then, the conclusion must be that recognition scores are reasonably good measures of advertising exposure and perception, but probably not good measures of attitude or behavioral change. Recall scores are not reliable either as measures of exposure and perception nor as measures of attitude and behavioral effects. However, both kinds of measures are widely used, possibly due to their ready availability from syndicated services, a reluctance to change habitual patterns, and the difficulty of obtaining more suitable measures of effects.

Attitudes and Behavior Toward Product

The previous sections on post-testing have dealt with the direct measurement of attention to and remembrance of the ads themselves. The other major category of post-testing deals with the effects of the advertising on attitudes and behavior toward the product advertised.

Research design. A consideration of research design must enter here. Although previously mentioned in the section on the Research Process, some elaboration will clarify the requirements for measuring effects of an ad on product attitudes and behavior.

One can measure attitudes toward a product after the advertising has appeared (after-only design), or both before and after the advertising in order to detect change (before-and-after design), or before and after the advertising among both an exposed and an unexposed group (before-and-after-with-control design), or after advertising with both an exposed and unexposed group (after-only-with-control design).

These designs may be shown graphically as follows, with A representing the appearance of advertising, and M representing some form of measurement:

	Before		After
After-only	--	A	M
Before-and-after	M	A	M
Before-and-after-with-control			
Exposed group	M	A	M
Unexposed	M	--	M
After-only-with-control			
Exposed	--	A	M
Unexposed	--	--	M

Only the latter two designs--those using an unexposed control group--provide valid measures of advertising effects.

This is a very important point; it can be illustrated with the following fictitious example, in which it is assumed that a campaign for Sparkle-Cola soft drink has appeared.

Design	Percent of respondents with favorable attitude to Sparkle-Cola				Interpretation & Questions
	Before		After	Change	
After-only	?	(ad)	25%	?	25% have the right attitude after the advertising. However, who knows if this is a change in attitude, or if so how much?
Before-and after	20%	(ad)	25%	+5	A 5% favorable change in attitudes occurred. However, who knows if this was due to the test ad or to other events that occurred at the same time--reduction of competitive ads, increased advertising of Sparkle-Cola in other media, the fact that the ad appeared on a hot day and attitudes toward all soft drinks improved, etc.?
Before-and-after-with-control					Plus 5% change in exposed group; plus 5% in non-exposed group; net change, zero. The ad had no effect on attitudes over and above other influences, for the reasons mentioned above.
EXPOSED	20%	(ad)	25%	+5	
NON-EXPOSED	15%	(--)	20%	+5	

Design	Percent of respondents with favorable attitude to Sparkle-Cola				Interpretation & Questions
	Before		After	Change	
After-only-with-control					A net difference of 5% between the exposed and non-exposed group could be real effect of ad if statistically significant
EXPOSED	?	(ad)	25%	?	
NON-EXPOSED	?	(--)	20%	?	

These fictitious data illustrate the point that only when a non-exposed control group is measured can one be sure that any observed change is due to the advertising. This merely emphasizes again the point made in the section on research design--cause and effect (where an advertisement is the independent "cause" variable) can only be determined through a controlled experiment. There are some other conditions necessary, of course, for complete control--a group of respondents must be randomly divided into two groups, and the advertising must be controllable so that only the test group has the opportunity for exposure. Further, its validity is increased if it is a controlled field experiment.

To re-emphasize it strongly--post-tests of advertising effectiveness in terms of attitude or behavior effects are not valid unless a control group is used. The after-only or before-and-after designs cannot reveal the effects of advertising.

Given that the control group condition is satisfied, however, post-test experiments of advertising effectiveness are generally concerned with measuring awareness, attitudes, and behavior in the form of sales inquiries or sales themselves.

Awareness. One objective of advertising is to create awareness of a new product, or increase salience in the case of an established one.

Previous research has shown that awareness--or "top-of-the-mind"--is highly related to actual sales. Survey questions to determine awareness are phrased about as follows:

> "When you think of toothpaste (or other product),
> which brand do you think of first?"
> "What other brands come to your mind?"

This may seem like a naive approach, but its simplicity is deceptive. Experience shows it to be one of the most valuable survey questions for measuring advertising effects, because it is directly related to actual sales.

Attitudes, beliefs, buying preference. In general usage, attitude toward a product can mean several things--like or dislike, belief (or image) and buying preference.

In precise meaning, however, *attitude* is "a favorable or unfavorable feeling toward the product"--the overall degree of liking or dislinking.

The *beliefs* about a product are sometimes called its reputation or "product image." Take a toothpaste brand, for example--it has several attributes including price, taste, decay preventiveness, and so on. It may rate high on one belief, low on another, and just average on another. A product's belief ratings may or may not be related to its attitude rating--for example, a customer may have an unfavorable attitude toward a toothpaste even though the price, taste, and decay-prevention properties are rated high. Why? Because the important factor, to that customer, is overcoming bad breath or some other factor.

Buying preference is another kind of verbal report, sometimes called "intention to buy" or "buying predisposition." This too can be independent of attitude and beliefs; a customer may have a favorable attitude and favorable beliefs toward an advertised brand, but have no intention of buying it. Why? Because he always buys groceries in a particular shopping center where his most favored brand isn't carried; rather than go out of his way, he always buys the chain store brand because it's more convenient--even though it may rate lower on brand attitude and other beliefs.

Survey questions on buying preference may be phrased as follows:

"If you were going to buy toothpaste tomorrow,
which brand would you be most likely to buy?"
"Which other brands would you consider?"
"Are there any brands you would definitely not
consider?"

The last question is sometimes included to determine the very important difference between a neutral reaction and an actively negative one.

Instead of individual questions on buying preference, a buying intention scale is sometimes used, with the respondent to check the one position on the scale that best expressed his intentions. Such a scale might take the following form, in the case of a particular brand of toothpaste:

"Check the one statement that best fits your
point of view:

_____ I'm definitely going to buy some right away.
_____ I'll probably buy some sometime.
_____ I might buy some in the future.
_____ I'll probably never buy any.
_____ I wouldn't use it even if it were given to me.

Buying intention scales can be useful if carefully developed (through research, not from an armchair) and validated against actual usage. If precise, the buying intentions of a group of consumers can be plotted on this "ladder of preference" and the effectiveness of advertising judged by the extent to which it moves consumers up the ladder.

Inquiries and sales. Another method of post-testing advertising effectiveness has already been described in the pre-testing discussion--

behavioral measures such as coupon returns or inquiries, and sales. Inquiries are direct responses to advertising--either through coupons or in the form of some contact with the advertiser or his retail outlets. Requests for more information, requests for samples or demonstrations, visits to retail outlets, etc., are more forceful indicators of effects than verbal reports. In short, they are forms of action or behavior requiring some effort on the part of the target audience. As in the case of awareness and the other verbal reports, one can be sure that these are direct effects of advertising only if the controlled experiment has been conducted.

Sales or purchases are, of course, the ultimate desired effect of advertising. One must distinguish between actual purchase and verbal reports of purchase. A survey question asking "Which brand do you usually buy?" or "Which brand did you purchase last?" will not give the same results as an actual sales measurement. Some products, such as cars and appliances, require that a warranty or registration card be filled out showing the purchaser's name at the time of sale, thus providing an unobtrusive measure of actual purchase. Most products, however, don't yield a direct sales record identifying the buyer, and indirect purchase measures must be used.

Store audits are often used to check the number of product units that have been sold. In the case of separate test markets for separate advertising treatments, it's possible to determine advertising effects.

Another indirect sales measure is the pantry inventory or household inventory. In visiting a sample of homes, an interviewer may ask to see the brands actually on hand in the household, along with the usual questionnaire interview. Again, the researcher must be able to link the possession of a particular brand of product with a controlled experiment design in order to determine the effect of advertising.

A measurement of behavior--inquiry, sale, store audit or product possession--is undoubtedly a better indicator of advertising effect than a verbal report of awareness, knowledge, attitude or preference. Because of problems in gathering the data, behavioral measures are not as widely used. But in some situations, behavioral measures may be easier and cheaper to obtain, and they should certainly be obtained if possible.

Analyzing Post-Test Results to Find Advertising Principles

Two major purposes of post-testing were mentioned earlier--one is to evaluate performance of particular ads or campaigns, the other is to determine some general principles that will be helpful in future advertising.

Combined ratings and content analysis. One way to arrive at general principles is to look at the content differences between ads which score high and those which score low on various readership measures. This procedure is a combination of content analysis and post-testing measurement.

For example, take the case of an advertiser who uses magazines extensively and gets Starch readership reports on the ads. He has placed 100 ads in a general magazine over a period of time and has readership scores on all of them. To look for general principles, he first separates the ads into three groups--25 scoring extremely high, 25 scoring extremely low, and 50 which have medium scores. What are the apparent content differences between the high-scoring ads and the low-scoring ads? A visual inspection may reveal the following:

	25 ads with High "% Noted"	25 ads with Low "% Noted"
Black-and-white	2	12
Two-color	8	8
Four-color	15	5
No illustrations	3	3
Small illustration	8	7
Large illustration	14	15
Question in headline	13	4
Direct statement	6	11
No headline	6	10

etc.

From this analysis, he can conclude that ad noted scores can be improved by using four-colcr and by using question headlines; use of an illustration apparently makes no difference. There are many other physical ad characteristics which can be analyzed in the same way, of course; different layout techniques, different persuasion strategies, and so on.

What measures are used in such an analysis? In addition to noted and other readership scores, high- and low-scoring ads can be compared for differences in attitude change, preference change, sales change--in short, any measure for which valid data are available.

Further analysis can be conducted among different vehicles of a medium. For example, the high- and low-scoring ads in a man's magazine may differ according to headline but not according to use of color; a woman's magazine may show the opposite tendency; in a specialized technical magazine, none of the layout differences may be important while persuasion strategy may be a crucial factor.

The number of ways in which such an analysis may proceed is limited only by the number of observable differences in high- and low-scoring ads, and the insight of the researcher doing the analysis. The principles arrived at through such analysis, however, should be checked out and verified through further usage, testing and analysis, or through experimentation.

This combination of content analysis and test scores has also been used in determining principles of television advertising. The Schwerin Corporation, for example, pre-tested thousands of TV commercials and analyzed them in this fashion to provide guidance for their clients. (Schwerin, you will remember, used before-and-after changes in "buying preference" verbal reports as the measure of commercial effectiveness.) They have discovered such things as the following:

> in cartoon commercials, animals are more effective than people as the main characters.
>
> Commercials containing a "family authority" figure (grandmother, older brother, etc.) are very effective among older viewers but not among young viewers.
>
> Humorous commercials are less effective than non-humorous commercials.
>
> Commercials that name the competition are more effective than those which don't.

Many other principles have been found through such analysis. It can be applied to any vehicle, any medium, using any measure of advertising effect.

Content analysis of competitive advertising. In order to know precisely what the competition is doing, it is sometimes desirable to analyze what they're saying. By collecting competitive ads or commercials, then quantitatively analyzing their approaches--in terms of themes, persuasion strategies, graphics, etc.,--a great deal can be learned. Other measures are needed, of course, to determine how effective the competitive approaches are, but a description of what they say is a necessary first step which is often neglected.

CHAPTER XI. MONEY RESEARCH: ADVERTISING EXPENDITURE

Money, or advertising expenditure, research deals with questions of how much money to spend, when to spend it and where to spend it.

How NOT To Do It

There are many methods of making wrong decisions on advertising expenditure. Three common erroneous procedures are "seat of the pants," "percentage of sales" and "watch the competitor."

The "seat of the pants" procedure goes something like this: the president and the principal officers of a company sit down around a long conference table and they discuss, debate and argue the question, "How much money should we spend on advertising next year?" All of them will have opinions on the matter--based on their own special interests, prejudices for or against the value of advertising, desire to impress people outside the company, and so on. Sometimes an entire marketing budget will have already been established and the discussion will center around what proportion of that total should go to advertising. After hours or days or sometimes months of wrangling--all based on opinions rather than evidence--some sort of compromise solution will be arrived at. This will be heavily weighted in the direction of the opinion held by the highest ranking company officials taking part and the most voluble, persistent and persuasive talkers at the table. The eventual figure will bear little or no relation to what really should be spent--it may be too much, in which case money will be wasted, or it may be too little, in which case extra profits will be lost.

"Watch the competitor" is a game based on the assumptions that the leading competitors are (a) smarter or (b) have some divine guidance or (c) have some real knowledge based on research. The advertising budget of the "watching" company is then geared to the budget of the competitor in some fashion. Chances are, however, that the competitors are also operating on faith and ignorance. Even if the competitors' advertising is based on research knowledge, what's right for one company isn't necessarily right for another; for example, advertising spending for a company that clearly leads the field in sales may be either too much or too little for companies with a smaller share of market.

The "percentage of sales" method gives the comfortable illusion that it is scientific because it is based on numbers and statistics. Sometime in the dim past, a company oracle may have decided that "advertising should be 3% of gross sales income." If this policy is followed year after year, it's possible to achieve a high correlation between sales and advertising. The fallacy is that cause-and-effect is reversed--instead of advertising being a cause of sales, it ends up being an effect of sales.

These fallacious methods are also used in determining which media to use and which messages are communicated. Here, however, we're concerned with the money question.

A Better Way: Controlled Field Experiments (CFX)

The most acceptable method for determining advertising expenditure is the controlled field experiment using sales (or profits) as the dependent variable. As with other advertising research methods, experts are not in complete agreement on that statement; the best support for it comes from some of the largest and most profitable advertisers in the country who have adopted it. It satisfies the businessman's requirement for practical and

and realistic goals, and the researcher's requirement for valid and rigorous measurement.

A complete description of controlled field experiments has already been given in the chapter on "common advertising research plans." (It may be useful at this point to review that section.) Some uses of CFX will be given here for several problems of advertising expenditure--total advertising budget, spending patterns over time, market concentration, media allocation, and vehicle allocation.

Setting Total Advertising Budget

Planning a controlled field experiment on the total advertising budget should start many months before a new budget is to actually go into effect. Why? Because advertising's effects are slow to take hold, except in special cases, and it is desirable that advertising be allowed to run for some time before making a final judgment.

Here is an example, again using the fictitious product, Sparkle-Dent toothpaste. The main dependent variable will be sales, to be determined by store audits. (With other products, sales might be determined by sales registration cards, pantry inventory or surveys of reported purchase.) It may be desirable to also measure such things as brand awareness, knowledge, image, attitude, buying preference, and so on. While a measure of sales alone will determine what happened, the other measures are needed to determine why they happened and give a better understanding of the dynamics of the advertising.

The experimental unit will be markets (metropolitan areas) rather than individuals or small geographic tracts. The reason for this is that it is next to impossible to manipulate several variations of spending level in all media within a market. A whole market can be controlled and varied more precisely.

Another fact of life is that it is usually impossible in the case of an established product to completely withdraw all advertising and set up a "no advertising" control treatment. In this example, the control treatment will be the normal advertising budget. Spending levels will be varied above and below the normal level to answer the questions, "Should we decrease or increase the advertising budget? If so, by how much?"

This example assumes that, from pre-research planning and discussion, management has agreed to test the following expenditure levels: one-half of normal (50%), normal (100%), one-half above normal (150%), and twice normal (200%). It's possible, of course, to test any other levels that are considered practical, all the way up to ten times normal or more if desired.

Another assumption in this example is that advertising will be varied proportionately in all media if possible. In the 50% treatment for example, where the advertising is cut in half, someone might note that television makes up half the total advertising budget and perhaps it would be easier to simply cut out television. This procedure would, however, confuse the effects obtained: it would be impossible to tell if an obtained change in sales was due to "budget decrease" or to television only, or if similar results would be obtained with decreases in other media.

Another assumption in this example is that the advertising messages themselves will be handled "normally" in all treatments. If both spending level and advertisements themselves were varied, it would be impossible to tell which factor caused any obtained change.

In short, in an experiment on spending level alone, all other conditions (media used, messages, etc.) should be held constant.

Having made those decisions and assumptions the details of the spending experiment can now be planned, with the following ingredients specified:

Product (fictitious):	Sparkle-Dent Toothpaste
Independent variable:	Spending level
No. of treatments:	Four (50% normal, 100% normal, 150% normal, and 200% normal)
Primary dependent variable:	Sales (data gathered by store audits)
Other dependent variables:	Awareness, knowledge, image, attitude, buying intentions (data gathered by personal interview sample survey)
Advertising timetable:	January-December (12 months)
Research timetable:	Sales: continuing measurement for six months before, during, and six months after. Surveys: before, during and after test.
Measurement units:	40 markets, randomly divided into four groups of 10.

At this point, as in other advertising research projects, it is useful to set up a time diagram in this manner:

July 1 Jan. 1 Jan. 1 June 30

Treatment	Pre-test Period (6 mos.)	Test Period (12 mos.)	After-test Period (6 mos.)
A (10 markets)	$: 100%	50%	100%
B (10 markets)	$: 100%	100%	100%
C (10 markets)	$: 100%	150%	100%
D (10 markets)	$: 100%	200%	100%
Sales audit	Monthly,	continuous--------------	
Surveys	x Dec.	x x x Mar. June Sept.	x x Jan. June

Note that monthly sales figures will be accumulated, permitting detection of when the sales changes occur and how long after the test period they persist. For the same reason, surveys will be conducted just before the test, at three-month intervals during the test, immediately after and six months after. For other products, these time intervals might vary, depending on marketing conditions, the natural purchase cycle and other factors.

From here on, there are many other details to be planned--the advertising agency must arrange to buy the media in the planned amounts, the sales auditing and survey procedures must be scheduled, and many other coordination efforts must be carried out among the company, the agency, the research firm and the media.

During the course of the experiment, rigorous attention must be paid to detailed execution of the plan. Analysis of the data is carried out throughout and after the experiment, resulting in a final report and a budget decision based on the results.

The bare bones of the analysis, however, are simple. The change in sales, attitudes, etc., for each treatment are compared with each other, and net changes compared with the control are computed. In outline, the analysis is as shown in earlier chapters:

Treatment	Measurement	Change (after minus before)
A (50%)	Before and after advertising	Amount, + or -
B (100%)	Same	Same
C (150%)	Same	Same
D (200%)	Same	Same

As noted before, advertising expenditure (the independent treatment variable) can be varied in many ways:

Thus: 50%, 100%, 150%, 200% (as in the example)
or: 0%, 100%, 200%.........1000%
or: 75%, 100%, 125%

It's worthwhile to reiterate one requirement for such experiments: all other conditions, except the independent treatment variable, in the test markets should be maintained exactly as normal. The money taken away or added to advertising should not be funneled into other marketing activities in those markets; local advertising by retail merchants should not be changed to compensate for the budget changes. Otherwise, the results of the experiment will be clouded. Regarding competitive advertising, there is of course no way to control this; one reason for using 10 markets for each treatment is to insure that fluctuations in competitive advertising and other non-controllable factors are constant for all treatments. The time, place, and conditions of such testing are generally guarded as "top secret" information, to prevent others within and without the company from fouling up the results, either deliberately or accidentally.

Here is a fictitious table of results that might be obtained from the example given:

	Sales Index Performance						
		After yr.-long test			Sales Change		
Treatment	Before	1 mo.	3 mo.	6 mo.	1 mo.	3 mo.	6 mo.
A (50%)	100	60	70	80	-40	-30	-20
B (100%-normal)	100	100	100	100	0	0	0
C (150%)	100	110	105	100	+10	+ 5	0
D((200%)	100	160	140	120	+60	+40	+20

The researcher can draw the following conclusions from these results: (1) changes in advertising budget have on immediate effect which gradually decreases as the budget is returned to normal; (2) sales are lost when advertising is decreased 50%; (3) sales do not increase appreciably when advertising is increased 50%; (4) doubling the budget produces a significant increase in sales which decreases somewhat over time but is maintained significantly higher as long as 6 months after the change.

The rate of decay in increased sales for the doubled budget could be determined by continuing the analysis for a year or more. One could also trace the change in profits as well as in sales; even though sales increased significantlv when advertising was doubled, it's conceivable that the cost per incremental sale was excessive, resulting in decreased profits along with increased sales.

Other analyses could compare the sales change with changes in awareness, knowledge, attitudes and other dependent variables measured.

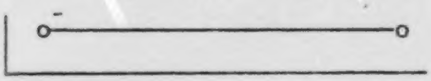
Determining the Advertising Spending Pattern

This refers to the timing of advertising expenditures over time. Given a certain amount of money to spend over a year's period--say, one million dollars--an infinite number of spending patterns is possible. A few are shown below:

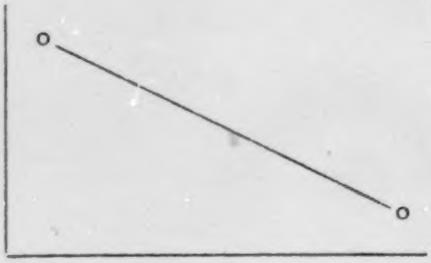
TREATMENT (PATTERNS) Spending by Months

1 2 3 4 5 6 7 8 9 10 11 12

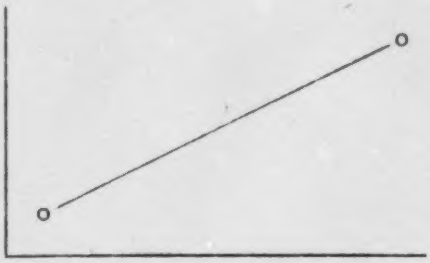
Rectangular: equal spending over time periods



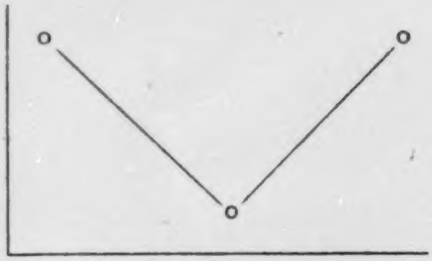
Downhill pattern: heavy spending at first, gradually decreasing to zero



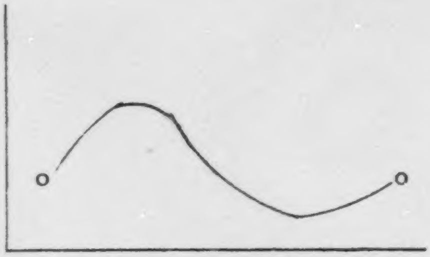
Uphill pattern: little spending at first, gradually increasing to maximum



V-shaped pattern: heavy at beginning and end of period, little at mid-point



Cyclical pattern: regular fluctuations based on other market or buying conditions



There are, of course, innumerable variations in spending pattern that might be tested--inverted-V patterns, W-shaped patterns, irregular patterns based on seasonal fluctuation, and so on. The correct pattern for one product may not be appropriate for another. Some products have seasonal usage and the variations in spending pattern might be geared to those fluctuations. Other products, such as automobiles, have annual product changes and the advertising expenditure may be geared to introduction time.

For the advertising researcher, the major point is this: any variations in spending pattern may be tested by controlled field experiments in much the same manner as budget experiments. From a pool of markets, markets can be randomly assigned to as many groups as there are spending pattern treatments to be tested.

The pattern of spending can be an extremely important factor in the results obtained from advertising. It's conceivable that a given advertising budget might result in 50% fewer sales or 50% more sales, depending on the spending pattern.

Determining Market Allocation of Funds

For one reason or another, it might be desirable to test whether a given advertising budget gets better results in some kinds of markets than others.

Some of the variations in market allocation that might be considered are these:

Market size:	TREATMENTS
	Large markets (over 500,000 population) vs. Medium-sized markets (50,000-500,000) vs. Small markets (under 50,000)

Market strength: TREATMENTS
Strong markets (where existing share of sales is high)
vs.
Average markets (sales share is medium)
vs.
Weak markets (sales share is low)

Geographical area: TREATMENTS
Far west
vs.
Midwest
vs.
East
(or a comparison of the seven major census regions)

Socio-economic conditions: TREATMENTS
High income (high per capita income & spending)
vs.
Medium income
vs.
Low income

There are many other ways in which different geographical locations, or markets, can vary. The important market characteristics vary from product to product, so no one style of market allocation is ideal for all products. Intensive study of other data on the product and its customers provide clues to the crucial experimental treatments.

The controlled field experiment can test these different strategies of advertising spending by allocating the same amount of money to several market conditions, and measuring the results.

Media Allocation Problems

Apart from the question of overall advertising expenditure, one problem that frequently arises in advertising is: Which advertising medium, or combination of media, would be best for placement of additional funds--radio, TV, newspapers, outdoor, direct mail, or some combination of those media?

In the chapter on media research, this problem was discussed in terms of such dependent variables as circulation, audience size, and advertising exposure; by such considerations, it's possible to narrow down the tremendous number of alternatives to a few alternatives for final experimentation. The controlled field experiment, with sales as the dependent variable, is really the acid test of media effectiveness.

To test the relative effectiveness of incremental spending in one or other of several media, the researcher conducts single media experiments. To test the relative effectiveness of combinations of media, the researcher conducts media mix experiments. As an example, assume there is $100,000 to invest in added media--that is, $100,000 over and above the normal budget.

Treatments

Single-media experiment: Normal media usage (control)
vs.
Normal plus $100,000 radio
vs.
Normal plus $100,000 TV
vs.
Normal plus $100,000 newspaper
vs.
Normal plus $100,000 magazine
vs.
Normal plus $100,000 direct mail

This example contains six treatments--normal usage to serve as a control, plus five treatments consisting of $100,000 extra in each of five media. It's possible, of course, to test only one additional medium in this fashion, or as many as are available, including outdoor, skywriting, matchbook covers, or whatever. In this case, with six treatments, the researchers would start with an initial selection of 12, 18, 24, 30 or any number of markets divisible by six--the more the better--and then randomly divide them into six treatment groups as described earlier. The results of this experiment would provide direct comparison of every medium with every other in terms of sales and profits, and/or various survey measures of effectiveness.

A media mix experiment is somewhat more complicated, since it involves a simultaneous test of combinations of media at various levels. To reduce the complexity, the example below will deal only with two media--newspapers and television--and two levels of additional spending in each. This produces four treatments as follows:

Constant sum media mix: Treatment

Normal (for control)
vs.
Normal plus $50,000 TV plus $50,000 NP
vs.
Normal plus $100,000 TV
vs.
Normal plus $100,000 Newspaper

The term "constant sum" above simply means that some fixed amount--in this case, $100,000--is allocated among two media. To determine the best constant-sum combination of six media could require 64 different treatments; in usual practice, owing to limitations in the number of test markets available, this number of treatments would not be feasible.

Variable sum media mix:

Treatment

1. Normal (for control)
   vs.
2. Normal plus \$50M TV
   vs.
3. Normal plus \$100M TV
   vs.
4. Normal plus \$50M Newspaper
   vs.
5. Normal plus \$100M Newspaper
   vs.
6. Normal plus \$50M TV and \$50M Newspaper
   vs.
7. Normal plus \$50M TV and \$100M Newspaper
   vs.
8. Normal plus \$100M TV and \$50M Newspaper
   vs.
9. Normal plus \$100M TV and \$100M Newspaper

In this variable-sum example, there are two media (newspapers and television) at three different levels--(zero, \$50 thousand, and \$100 thousand) for each, for a total of nine different treatments. This is a variable sum media mix because, even though no single medium gets more than \$100,000, the total spending per treatment varies from \$50,000 to \$200,000. The results of such an experiment can determine not only which individual medium is best in comparison with the other, but also the effects of different combinations of media. The design shown above was actually used by a major advertiser not too long ago. Another company conducted an even more complex study in which four different media, at two levels each, were tested--a total of 16 treatments. Depending on the number of test markets available and suitable, an advertiser could theoretically test any number of media, at any number of levels; in practice, 16 treatments becomes unwieldy to manage and is extremely expensive.

Vehicle Allocation Problems

Assuming it has already been decided to spend \$100,000 in the magazine medium, the next question is which vehicle to use. This would lead to a vehicle allocation study. Should the \$100,000 be spent in *Cosmopolitan*, *Redbook* or *American Home*? In the case of magazines, the researcher could depart from the use of markets as the measurement units, and instead resort to *split-run* experimentation, in which individuals are the measurement units. In the split-run technique, a listing of *individuals* is randomly divided into treatment groups; this requires that sales data be keyed to individuals rather than markets; this is more feasible for some products than others. In this case, let's say that individual sales information is available (through compare orders, for example).

Single vehicle experiment: *Treatment*

Normal spending (for control)
vs.
Normal plus $100,000 in Cosmopolitan
vs.
Normal plus $100,000 in Redbook
vs.
Normal plus $100,000 in American Home

The researcher could also *conduct constant sum vehicle mix* and *variable sum vehicle mix* experiments in the same manner described previously for media experiments.

Other Advertising Expenditure Problems

This chapter has described only a few applications of controlled field experiments to problems of total advertising budget, spending patterns over time, market concentration, media allocation and vehicle allocation. It's also possible to design experiments which would test various *combinations* of those allocations simultaneously--for example, a simultaneous test of total advertising budget levels and various spending patterns. Such complex studies are relatively easy to design, but they are difficult to execute. Several reasons for this difficulty are limited numbers of suitable test markets, a high number of possible treatment combinations, the attention necessary for executing the details of the design, and coordination of the efforts of company, advertising agency and media. One rule of thumb is that a controlled field experiment becomes difficult to manage when more than 12 treatments are involved. This is not to say that a large number of treatments is impossible, but only that the problems multiply much more rapidly than the number of treatments.

The controlled field experiment, with sales or profits as the dependent variable, embodies the best business and scientific approaches in advertising research today. The plan is not suitable for all problems concerning media or messages or money--but where possible, it is the method of choice.

There is a strong belief by many that statistical analysis of data, without conducting any additional or new research, can provide answers to many advertising problems. There are equally strong beliefs by many experts that statistical analysis alone is an attempt to make "a silk purse from a sow's ear," meaning that the statistical approach used by itself is purely wishful thinking. Some of those statistical techniques are discussed briefly in the next chapter.

CHAPTER XII. MISCELLANEOUS TOOLS AND TECHNIQUES

The research methods, plans, tools and techniques already described probably account for 95% of the time and effort being expended on advertising research. Out in front of this wave, however, are the advanced thinkers experimenting with other scientific approaches.

Just as in earlier examples, it's possible to understand some of the concepts and the applications of these tools without understanding all the inner workings, technical details and mathematical proofs.

Some of these innovations are described very briefly below. The serious advertising researcher will keep in touch with these and other new developments by reading the publications already mentioned in an earlier chapter. Some of the most pertinent are the Journal of Advertising Research, Journal of Marketing and Journal of Marketing Research.

Operations Research

Operations research is the umbrella label given a number of specific mathematical techniques that have been helpful in making advertising, marketing and other business decisions. Computers are generally, but not always, used for the complicated statistical manipulations involved. Included in operations research are some of the other tools described below. It has been very useful in making practical decisions about inventories, prices, production scheduling and transportation scheduling.

Statistical Decision Theory

Statistical decision theory is one application of operations research The problem to be solved is generally one of choosing between two or more alternative courses of action. SDT provides estimates of the quantitative "value" of the different outcomes, and the likelihood that each will occur under a given set of specified conditions. For example, two alternative courses of action for next year's advertising expenditure might be "twice as much as this year" and "half as much as this year." The different outcomes are measured only as "increased sales" or "decreased sales." (Independent variable, advertising expenditure; dependent variable, change in sales.) The basic question: what are the respective probabilities that (a) more sales will be obtained with more advertising, (b) more sales with less advertising, (c) less sales with more advertising, and (d) less sales with less advertising? SDT allows a choice between more advertising and less advertising on the basis of the highest net value of the combined outcomes (more sales and less sales) of each.

Mathematical Models

A mathematical model is an equation expressing the relationship between a number of independent variables and a dependent variable. Take the dependent variable, "number of sales" (X). The things that affect sales (independent variables) might be "number of salesmen" (A), "number of retail outlets" (B), and "number of advertising pages" (C). Then a value must be assigned to each independent variable unit. Let's say it is estimated that a retail outlet is 1000 times as valuable as one advertising page, and a salesman is 100 times as valuable as an advertising page. The equation might end up like this:

$$1000A + 100B + 1C = X$$

In short, a mathematical model is a numerical expression of the factors that theoretically influence some dependent variable.

Simulation

Simulation involves the use of mathematical models to predict artifically "what would happen if . . .". Using the above model (with values or "weights" for each of the independent variables arrived at through previous experimentation), one can simulate a number of different situations and choose the one that comes out best. For example, assume a company has an extra million dollars to spend on marketing next year. Should it be spent on new salesmen, new retail outlets or more advertising--or some combination of all three? One million dollars will allow the addition of either five retail outlets or 30 salesmen or 20 advertising pages. Sales for each of these alternatives are simulated below:

Alternatives: Invest one million dollars in . . .	Equation
Five retail outlets	1000 (5) + 100 (0) + 1 (0) = 5,000
30 sales men	1000 (0) + 100 (30) + 1 (0) = 3,000
20 advertising pages	1000 (0) + 100 (0) + 1 (20) = 20

Through simulation, the conclusion is drawn that the extra money would best be spent on retail outlets. (The values of the different units above are fictitious, not based on real data, simply to illustrate the procedure.) As can be noted in the above example, the crucial number for the accuracy of simulation is the weight or value assigned to one unit of each independent variable; if that isn't accurate, the equation is worthless and can be quite misleading. Simulation is of great value for manipulation of physical objects, of questionable validity for determining effects on human thoughts and behavior.

Linear Programming

Linear programming can be illustrated as an application of simulation to the choice of media combinations or media mixes. For example, suppose a company has one million dollars to spend on magazine advertising and they want to divide it among five magazines. The problem is, there are 100 different magazines available and over 75 million different combinations of five magazines that might be chosen from the list of 100. Each combination of five magazines represents one equation--thus, over 75,000,000 equations must be computed. Linear programming via computer makes it feasible to simulate all the different "mixes" and choose the one that will apparently maximize reach for a given amount of money.

Experimental Design

Various applications of experimental design already have been described under various labels in preceding sections of the book--laboratory experiments, controlled field experiments, split-run testing and so on. The classic experimental design involves two groups--a test group and a control group, assigned by randomization of the total pool of subjects--with after-only measurement, or before-and-after measurement; only the test group is exposed to the treatment variable.

More complex experimental designs are: factorial designs, where several variations of two or more independent variables are tested simultaneously; the latin-square design; the randomized-block design; and many others.

One crucial point to remember about experimental design is that it is an essential condition for determining cause and effect. It differs from simulation: in simulation, one tries to predict what would happen in various situations by artificial manipulation of numbers; in experimentation, the various situations are actually tried out in the real world.

Needless to say, experimentation is many, many times more expensive--but the value of its conclusions is also much greater.

Factor Analysis

Factor analysis is a statistical technique used to reduce a great many variables down to a few general factors underlying them.

"Image" studies provide a good example. The image of a car is composed of many beliefs about styling, interior appearance, riding comfort, durability, initial cost, cost of operation, pickup and getaway, trade-in value, and hundreds of other product attributes on traits. In trying to determine which of these are most important, it may be found that some of the beliefs are almost perfectly correlated with each other. For example, styling and interior appearance beliefs correlate to form an "attractiveness" factor; initial cost, cost of operation and trade-in value may all be perfectly correlated with each other to form an "economy" factor; durability and pickup might go together to form a "strength" factor.

After a factor analysis of all the beliefs that make up an image, one may end up with only three or four basic factors; future image measurement could then concentrate on these factors instead of measuring 100 different beliefs.

These "factors"--e.g., attractiveness, economy, strength--are more likely to be significantly related to sales than the numerous original variables.

CHAPTER XIII. WHAT THE PUBLIC THINKS ABOUT ADVERTISING

Most of this book on advertising research deals with measuring advertising effects for specific products. Here is quite a different kind of research study--one that attempted to find out what the public thought about advertising in general.

The members of the American Association of Advertising Agencies (AAAA) had been concerned for some time about various criticisms of advertising in general. Some of the familiar criticisms are that advertising is offensive or irritating, or that there's too much of it, or that it raises product costs, and so on.

Therefore, in 1964 the AAAA decided to take a scientific approach and find out just what the public really did think, instead of relying on the scattered opinions of a few writers and critics. The purpose was to locate and understand any widespread criticism, if any, of advertising on the part of the public. There were two main parts to the study: one, public reactions to advertising as an institution; two, public reactions to advertisements themselves.

The basic research procedure was a national survey of a random sample of the "consumer population"--persons 18 and over. Two personal interviews were conducted with the members of the sample. In the first interview, a standard question-and-answer procedure was followed to determine attitudes toward advertising as an institution. The second part of the data-gathering process had some unusual features and will be described after presenting the results of the first part below.

Attitudes Toward Advertising as an Institution

This part of the research was intended to answer these questions:

> How important is advertising to the American public, in comparison with a number of other institutions?
>
> How favorable or unfavorable does the public feel toward advertising in general?
>
> How does the public feel about certain specific elements of advertising--economic aspects, social aspects, the content, and so on?

Many questions were necessary to get specific answers to these questions, but the principal results were as follows:

Importance of advertising. Advertising is not a central issue in the day-to-day thinking of the American public. People are more immediately concerned with other aspects of life such as raising children, religion, education and the government.

- In a list of 10 different aspects of life, advertising ranked last as a topic of conversation; only 11% said it was a subject they talked about most.
- Advertising ranked last as something they have strong opinions about, with only 7% reporting strong opinions on it.
- Advertising ranked third among the ten institutions as something they enjoyed complaining about without being too serious in their complaints, with 20% giving that response.
- Advertising ranked fifth as needing immediate attention and change, with 15% feeling that way.

The table below summarizes the public's opinions on the importance of advertising in comparison with other aspects of life.

	Percent who reported each topic as one they...			
	Talk about the most	Have strong opinions on	Enjoy complaining about but not really serious about	Feels needs immediate changes
Bringing up children	58%	43%	12%	20%
Family life	49	35	14	12
Religion	47	52	6	10
Public education	41	33	14	41
Federal government	36	32	33	28
Clothing and fashions	35	12	23	6
Professional sports	33	11	11	5
Labor unions	18	13	17	26
Big business	13	8	16	11
ADVERTISING	11	7	20	15

By any standards, advertising was not felt by the American public to be a very important aspect of their lives.

Attitudes toward advertising. Three times as many people had favorable attitudes toward advertising as had unfavorable attitudes. More specifically, 41% were favorable, 14% unfavorable and the remaining 45% had mixed or indifferent attitudes.

Those who were favorable toward advertising felt that way primarily because it was "informative" and "good for the business economy." Those who were unfavorable felt that way primarily because it was "false or misleading," "there's too much of it," and "it interrupts entertainment."

Attitudes toward specific elements of advertising. The public was discriminating about some good and bad features of advertising. On the one hand, most agreed that advertising is essential (78%), that it helps raise our standard of living (71%), and results in better products for the public (74%). On the other hand, many thought it persuades people to buy things they shouldn't (65%), relatively few though it results in lower prices (40%), and about half thought it doesn't present a true picture of the product (53%).

At the same time, more than half the public believed that today's standards of advertising are higher than 10 years ago; 58% said higher, 10% said lower and the rest gave other responses.

In summary, it is clear that the public was quite able to point out certain aspects of advertising which are regarded highly, and others which are criticized. It was seen as neither black nor white but, like other aspects of life, as having some good and some bad features.

Reactions to Specific Advertisements

For the second part of this study of public reaction to advertising, every person in the study was given a small hand counter and asked to click it each time he paid any attention to an ad in magazines, newspapers, radio or television. In addition, he was asked to comment in a small notebook on any of the ads that he found particularly good (enjoyable or informative) or particularly bad (annoying or offensive). This procedure led to information on how many ads were consciously noticed by a person on an average day, how many of those he had some strong positive or negative reaction to, and other information about the kinds of ads that produced these reactions.

Number of ads perceived. The average person noticed about 76 ads per day (or at least noticed them enough to click his counter when exposed). This represents only a small fraction of the physical opportunities for exposure to ads in the average day's intake of newspapers, magazines, radio and television. No one knows exactly how many actual exposures this represents, but it undoubtedly runs to several hundred, and one estimate several years ago was 1500 advertising exposures per day per person. Seventy-six advertising perceptions out of 1500 exposures, if anywhere near accurate, means that a person notices (perceives) only about 5% of the ads physically encountered (are exposed to), or that about 95% of all advertisements can have no effect.

Women tended to perceive slightly more ads per day than men (80 to 70), people aged 35-49 were slightly above average in number perceived (84) and persons over 65 years of age perceived somewhat fewer ads (62). But, in general, most kinds of people showed few differences in advertisement perception.

Number of ads creating strong reactions. Out of the 76 ads perceived per day, about 64 created no strong reactions one way or the other, while about 12 (or 16%) did evoke a positive or negative comment. To summarize up to this point:

> The average person on the average day
> ... is *exposed* to about 1500 ads (previous estimate)
> ... *perceives* about 76 of those,
> ... *reacts strongly* to about 12 of those, of which about 8.6 are favorable reactions and 3.4 are unfavorable reactions.

What are the characteristics of that small proportion of ads which evoked strong reactions either positive or negative?

Among the "strong reaction" responses, 23% were annoying, 5% were offensive, 36% were informative and 36% were enjoyable. That is, when there is any conscious reaction to an ad at all, it is more than twice as likely to be favorable as unfavorable.

However, as might be suspected, certain kinds of advertisements are more likely to be annoying or offensive than others.

Out of 78 product categories mentioned, five received predominantly negative reactions. Against the average 5% "offensiveness index" for all ads, liquor ads scored 55% and cigarets scored 19%.

Either because of the nature of the product, or the nature of the advertising, those product-groups were more annoying or offensive than others.

Summary

This AAAA study went on to show why certain ads were seen as informative, enjoyable, annoying and offensive, and such research data might be useful in making advertising more palatable to the public. The general conclusions were as follows:

(1) Advertising is not seen by most of the public as being a topic of great concern compared with other aspects of daily life.

(2) More people have favorable than unfavorable attitudes toward advertising in general.

(3) There are no great differences among different kinds of people in their reaction to advertising.

(4) People consciously notice only a very small proportion of the ads to which they're exposed.

(5) A substantial majority of the ads noticed don't stimulate any strong feelings, either positive or negative.

(6) Among that small proportion of ads which do evoke a strong reaction, the favorable reactions outnumber the negative ones.

(7) Some kinds of products are more likely to get negative responses than others.

This broad look at public attitudes toward advertising in general may be useful in helping research do a better job for advertising.

-- END --